# GENERALIZED POLYKAYS, AN EXTENSION OF SIMPLE POLYKAYS AND BIPOLYKAYS[1]

By Eugene Dayhoff

*Texas A & M University*

**1. Introduction and summary.** In an earlier paper [1] the author presented a generalization of the second degree bipolykays of Hooke [2], defined for arbitrary balanced population structures, and showed the equivalence of these generalized polykays and the $\Sigma$ functions defined by Zyskind [7].

In this paper is presented a more general formalization of generalized symmetric means and polykays of arbitrary degree and some sampling properties of these. Utilizing the fact that the second degree generalized polykays are equivalent to the $\Sigma$'s, which are defined in terms of components of variation, an application to obtaining the variances of estimates components of variation is also presented.

**2. Review of some basic concepts.** The concepts presented in this section are for the most part extracted from [2] and presented here for convenience of the reader.

Tukey [4] denoted symmetric means and polykays by brackets and parentheses, respectively. For example $\langle ab\rangle = \sum^{\neq} x_i^a x_j^b/N(N-1)$, where the sum is over all subscripts with the restriction that the different subscripts must remain unequal. The corresponding polykay would be denoted $(ab)$. Hooke [2] introduced a secondary notation for these functions. The notation for $\langle abc\rangle$ for example would be $\langle p_1p_2 \cdots p_a,\ q_1q_2 \cdots q_b,\ r_1r_2 \cdots r_c\rangle$, where $p_i$, $i = 1, \cdots, a$, denotes the individual $x_i$ and $q_i$, $i = 1, \cdots, b$, the individual $x_j$, etc. The entries $a$, $b$, $c$ in the angle bracket are said to form a *partition* of the integer $m = a + b + c$, where $m$ is the degree of the symmetric mean. The comma in the secondary notation then, separates the *parts* of the partition with $a$, $b$, $c$ denoting the length of the partitions. A *subpartition* $\beta$ of a partitition $\alpha$, say, may be formed by inserting one or more commas between the letters of $\alpha$. Two partitions are said to be *equivalent* or not *distinct* if they are identical, except possibly for the order of parts and the order of symbols within a part. Hooke defined the simple polykays of Tukey, [4 ], [5], and [6], in the secondary notation as follows:

$$\langle\alpha\rangle = (\alpha) + \textstyle\sum_i (\beta_i),$$

where there is an equation for each distinct symmetric mean of the same degree and the sum is over all distinct subpartitions $\beta_i$ of the partition $\alpha$ (two symmetric

---

Received 13 August 1965.

[1] This work was done for the most part at Iowa State University with support by the Aeronautical Research Laboratory, Air Force Research Division under contract AF33(616)-8269.

means, or polykays, are equivalent, or not distinct, if the partitions representing them can be made equivalent by renaming the symbols).

Thus, for example, the simple polykays of degree three

$$\begin{aligned} k_{111} &= \langle 111\rangle, \\ k_{12} &= \langle 12\rangle - \langle 111\rangle, \\ k_3 &= \langle 3\rangle - 3\langle 12\rangle + 2\langle 111\rangle, \end{aligned}$$

are defined, in the secondary notation, by the equations

$$\begin{aligned} \langle p, q, r\rangle &= (p, q, r), \\ \langle p, qr\rangle &= (p, qr) + (p, q, r), \\ \langle pqr\rangle &= (pqr) + (p, qr) + (q, pr) + (r, pq) + (p, q, r), \end{aligned}$$

which may be solved to yield the following

$$\begin{aligned} (p, q, r) &= \langle p, q, r\rangle, \\ (p, qr) &= \langle p,qr\rangle - \langle p, q, r\rangle, \\ (pqr) &= \langle pqr\rangle - \langle p, qr\rangle - \langle q, pr\rangle - \langle r, pq\rangle + 2\langle p, q, r\rangle. \end{aligned}$$

In developing the bipolykays, Hooke supposed a population matrix $\|x_{IJ}\|$, $I = 1, 2, \cdots, R$; $J = 1, 2, \cdots, C$, from which a bisample $\|x_{ij}\|$, $i = 1, 2, \cdots, r$; $j = 1, 2, \cdots, c$, is selected randomly. He defined generalized symmetric means to be average of monomial functions over a matrix, i.e., a gsm is a polynomial

$$M^{-1} \sum\nolimits^{\neq}_{p,q,\cdots,s,t} x_{pq}^{a_{pq}} \cdots x_{st}^{a_{st}},$$

where the symbol $\sum^{\neq}$, for a two-way array, means summation over all subsequent subscripts, with the restriction that row subscripts represented by different letters must remain different, and likewise for column subcsripts, and $M$ is the number of terms in the sum. The general term $x_{pq}^{a_{pq}} \cdots x_{st}^{a_{st}}$ contains $m$ fractors, of which $a_{pq}$ are equal to $x_{pq}$, etc., the degree of the symmetric mean being $m = a_{pq} + \cdots + a_{st}$. To each factor a different symbol is assigned and the resulting set of symbols partitioned in two ways—once by rows and once by columns. Thus the secondary notation for the gsm is an ordered pair $\langle\alpha_1/\alpha_2\rangle$ of partitions $\alpha_1$ and $\alpha_2$, each on the same set of symbols.

Hooke then defined a non-commutative "dot-multiplication" for symmetric means from a two-way population structure as follows:

$$\begin{aligned} \langle\alpha_1\rangle\cdot\langle\alpha_2\rangle &= \langle\alpha_1/\alpha_2\rangle, && \text{if } \alpha_1 \text{ and } \alpha_2 \text{ consist of the same set of symbols,} \\ &= 0, && \text{otherwise.} \end{aligned}$$

Distributivity is assumed to provide dot-products for linear combinations of symmetric means. The bipolykays are defined as

$$(\alpha_1/\alpha_2) = (\alpha_1)\cdot(\alpha_2)$$

where it is understood that the polykays $(\alpha_1)$ and $(\alpha_2)$ are first expressed as linear functions of symmetric means before the dot-multiplication is performed.

As an example,

$$\begin{aligned}(pq/p, q) &= (pq)\cdot(p, q)\\ &= [\langle pq\rangle - \langle p, q\rangle]\cdot\langle p, q\rangle\\ &= \langle pq/p, q\rangle - \langle p, q/p, q\rangle.\end{aligned}$$

**3. Extension of generalized symmetric means.** In this development it will be assumed the sampling involved is random and balanced so that the sample structure will be of the same form as the population structure. It is also assumed that a linear model can be written describing the kind of structure involved and containing a component for every factor in the structure. A typical observation in the population model or "identity" will be denoted by the capital letter $Y$, and will contain a different subscript, $I, J, K, \cdots$, for each of the factors in the population structure. These factors will simply be referred to as factors $A, B, C, \cdots$, and the population range of the subscripts will be as follows: $I$ a range of $A$ units, $J$ a range of $B$ units, and $K$ a range of $C$ units, etc. The corresponding sample observations will be denoted by small letters, the subscripts will be lower case letters and the range of the subscripts will be $a$ units, $b$ units, $c$ units, etc.

A factor represented by the subscript $J$, say, in a given response is said to be *hierarchal*, or *nested* within another set of factors whose representative subscripts belong to the subset $S_\alpha$ of the set of subscripts $S$ of the response, if the unique identification of each level of $J$ requires also the specification of the levels of each subscript of $S_\alpha$. A population, or sample, is said to be *balanced* with respect to all subscripts used in the representation of an arbitrary population or sample observation if the range of any one of the subscripts is the same for every set of particular values the other subscripts may assume.

We now proceed with a definition of a generalized symmetric mean of degree $r$ for an $n$-way crossed population structure.

DEFINITION 3.1. A generalized symmetric mean of degree $r$ from an $n$-way crossed population structure is a symmetric polynomial of the form

$$M^{-1} \sum\nolimits^{\neq} Y^{a_1}_{[1]} Y^{a_2}_{[2]} \cdots Y^{a_s}_{[s]},$$

where $\sum_{i=1}^{s} a_i = r$, the sum is over all possible selections of observations in the population subject to the requirements of the relationships of the subscripts, $M$ is the number of terms in the sum and

$$Y^{a_s}_{[s]} = [Y_{I^{\theta_1^s} J^{\theta_2^s} \cdots P^{\theta_n^s}}]^{a_s},$$

and the symbols $\theta_v{}^u$ are required to take any one of the form $0, 1, 2, \cdots, s - 1$, which will usually be the number of primes inserted to differentiate levels which must be different in the sum; that is, if $Y_{[u_1]}$ and $Y_{[u_2]}$ differ with regard to the $v$th factor then $\theta_v{}^{u_1}$ and $\theta_v{}^{u_2}$ must be different.

Let $W_v$ denote the number of different $\theta_v{}^u$'s. Then the number of possible terms in the sum with regard to the $v$th subscript is $N_v(N_v - 1) \cdots (N_v - W_v + 1)$, where $N_v$ is the population range of the $v$th subscript.

The gsm is symmetric in the sense that it is invariant under the permutation of any of the subscripts.

Since the constituent $Y$'s are allowed to be alike or different with respect to a given subscript, say $v$, $\theta_v^1, \theta_v^2, \cdots, \theta_v^s$ can be related in any manner ranging from being all alike to being all different. The only restriction in the sum is that the different $\theta_v^u$ must remain unequal. The $Y$'s are thus said to be *restricted* in the sum.

In an arbitrary structure we may classify the factors $A$, $B$, $C$, into two general groups, those that are nested in some of the other factors and those that are not nested in other factors. Denote a gsm from the $n$-way crossed structure by $h(\tilde{Y})$. Then for each gsm of the crossed structure a gsm for an arbitrary structure is defined as follows.

DEFINITION 3.2. For each gsm $h(\tilde{Y})$ from the $n$-way crossed population structure there exists a corresponding gsm, say $h(Y)$, for any balanced population structure which is defined as that symmetric mean which obtains when the following additional conditions are required of the subscript of $h(\tilde{Y})$:

(i) If the subscript corresponds to a factor which is not nested in any other factors, the $\theta_v^u$'s with respect to this subscript satisfy the same conditions as those of $h(\tilde{Y})$.

(ii) If the $v$th factor, say, is nested in the $v^*$th factor, and $\theta_{v^*}^{u_1} \neq \theta_{v^*}^{u_2}$ in $h(\tilde{Y})$, then $\theta_v^{u_1} \neq \theta_v^{u_2}$ must hold in $h(Y)$. In this sense subscripts which are nested by unequal subscripts are free to take values over their whole population range.

(In connection with this, the rule is adopted that a nested factor has different subscripts for every combination of levels of all nesting factors.) If $v$ is nested by $v^*$, $v^{**}$, etc., the above condition holds for each of the $v^*$, $v^{**}$, $\cdots$ independently. In condition (ii), the nested subscript is said to be *free*. In general there may be several groups of the $\tilde{Y}$'s of $h(\tilde{Y})$ in which all the subscripts in one group are alike with respect to the nesting subscripts but different from another group of like subscripts with regard to at least one of the nesting subscripts. The nested subscripts in this case are said to be *group-wise free* in the sum of the gsm.

Take as an example of these definitions the four factor completely crossed structure with factors $A$, $B$, $C$, $D$ represented by the subscripts $I$, $J$, $K$, $L$ in a typical response $Y_{IJKL}$, with ranges of $A$, $B$, $C$, $D$, units for the subscripts respectively. Then a typical gsm of degree four might be

$$(3.1) \quad \sum\nolimits^{\neq} Y_{IJKL} Y_{I'J'KL'} Y_{I'J'KL} Y_{I''J''KL'} / ABCD(A-1)(A-2)(B-1)(B-2)(D-1).$$

Suppose now that factors $B$ and $C$ are hierarchically arranged within factor $A$ but factor $D$ is crossed with the other factors. (3.1) then becomes, after imposing the additional restriction on the subscripts due to the nesting,

$$(3.2) \quad \sum\nolimits^{\neq} Y_{IJKL} Y_{I'J'K'L'} Y_{I'J'K'L} Y_{I''J''K''L'} / A^2B^2C^2D(A-1)(B-1)(C-1)(D-1).$$

**4. Alternate notation.** The secondary notation of Hooke's can be used to describe gsm's from $n$-way crossed structures but appears inadequate for gsm's from arbitrary structures. The secondary notation for (3.1) would be $\langle p,\ s,\ qr/p,\ s,\ qr/pqrs/pr,\ qs\rangle$. A similar notation for (3.2) would not adequately describe the nature of nesting. The following notation, not entirely unrelated to Hooke's, was found to be useful.

For any crossed population the letters $A$, $B$, $C$, $\cdots$, are used to denote subscripts which are primed alike, beginning with $A$ for no primes, $B$ for one prime, etc. (the use of these letters in this notation should not cause confusion with the usage in connection with identifying factors and the ranges of subscripts). A matrix is formed with as many rows as there are factors. If the degree of the gsm is $k$, then there will be $k$ letters in each row. The gsm in (3.1) would thus be represented as

$$\begin{vmatrix} A & B & B & C \\ A & B & B & C \\ A & A & A & A \\ A & B & A & B \end{vmatrix}. \tag{4.1}$$

This notation may be modified easily to indicate the nature of nesting and crossing in an arbitrary structure. As was mentioned earlier, the $Y$'s of the gsm may be grouped on the basis of the likeness of the combinations of all subscripts which nest a given subscript. The letters of the row corresponding to a nested subscript will be subscripted with numbers 1, 2, 3, $\cdots$, denoting the different groups of nesting subscripts. As before, the subscripts with a restricted range within a group will be indicated by use of the different letters $A$, $B$, $C$, $\cdots$, where $A$ will represent no primes, $B$ one prime, etc. Thus (3.2) would be written

$$\begin{vmatrix} A & B & B & C \\ A_1 & B_2 & B_2 & C_3 \\ A_1 & B_2 & B_2 & C_3 \\ A & B & A & B \end{vmatrix}. \tag{4.2}$$

These gsm's might also be conveniently written on a single line as, for example, $\langle ABBC/A_1B_2B_2C_3/A_1B_2B_2C_3/ABAB\rangle$. Parentheses would be used to denote a corresponding polykay.

Because the generalized symmetric means are invariant under permutation of the subscripts of the observations in the mean, this notation is not unique, that is to say, there may be several forms of the notation which represent a given gsm and the identity of each is often not easily seen unless certain permutations of the subscripts are made. For example, the gsm

$$\begin{vmatrix} A & C & B & B \\ A & C & B & B \\ A & A & A & A \\ B & A & B & A \end{vmatrix} \tag{4.3}$$

is just a permutation of (3.1), this readily being seen by interchanging the second and fourth columns and relabelling the letters in order from left to right. This method of checking for identity is equivalent to interchanging the parts of a partition and renaming symbols as discussed in Section 2.

**5. Random cross labelling.** The development in this section is introduced in order to define polykays from arbitrary structures more readily. The concept herein will be exploited also in establishing properties of the arbitrary polykays with reference to the corresponding properties of the crossed-structure polykays.

As was pointed out earlier the rule was adopted that a nested factor has different subscripts for every combination of levels of all its nesting factors. This convention is consistent with the physical situation implied in a nested population of individuals in that the different units of a nested factor in a particular combination of levels of nesting factors have, in general, no relationship to the units in another combination of levels of the nesting factors. On the other hand, if a factor is crossed with all other factors, the subscripts for this factor are the same, regardless of the combination of levels of the other factors. The physical implication here is obvious.

Through a process which shall be called *random cross labelling*, however, an "artificial" crossed structure may be formed from an arbitrary structure involving nesting. This process consists simply of randomly labelling the subscripts of a nested factor as subscripts of a crossed factor, the random labelling being done independently from one combination of levels of the nesting factors to another.

If a function of the observations of one of the possible "artificial" populations is formed, the average value of this function in terms of the observations of the arbitrary structure may be found by taking the expectation of the function over all possible ways of random cross labelling. Let $f(\tilde{P})$ denote a function of observations from the artificially induced population. Then the process of averaging $f(\tilde{P})$ over all possible random cross labelling shall be denoted by the symbol

$$E_{\text{lab}} f(\tilde{P}).$$

Specifically, gsm's of $\tilde{P}$ can be averaged in this manner to obtain gsm's of $P$ and thus provides a basis for defining polykays for an arbitrary structure with reference to polykays from the crossed structure as is shown in the next section. The following theorem relates the gsm's from the two structures:

THEOREM 5.1. *Let* $g(Y)$ *denote a generalized symmetric mean from the arbitrary structure* $P$ *which is the corresponding* gsm *of* $g(\tilde{Y})$, *a* gsm *from one of the artificially induced crossed populations, say* $\tilde{P}$, *as defined in Definition* 3.2. *Then*

$$E_{\text{lab}}\, g(\tilde{Y}) = g(Y).$$

PROOF. According to Definition 3.1, $g(\tilde{Y})$ is an average of terms of the form $\tilde{Y}_{[1]}^{a_1}\tilde{Y}_{[2]}^{a_2} \cdots \tilde{Y}_{[s]}^{a_s}$. The sum indicated in $g(\tilde{Y})$ is over both nested and non-nested subscripts. Let $\sum_0^{\neq}$ denote the sum over all the non-nested subscripts with the appropriate restrictions and $\sum_{n^*}^{\neq}$ the sum over all the nested subscripts of $g(\tilde{Y})$. Further let $N_0$ and $N_n^*$ denote the number of terms in the respective sums. Then

$$\begin{aligned} E_{\text{lab}}g(\tilde{Y}) &= E_{\text{lab}} \textstyle\sum_0^{\neq} \sum_{n^*}^{\neq} \tilde{Y}_{[1]}^{a_1} \tilde{Y}_{[2]}^{a_2} \cdots \tilde{Y}_{[s]}^{a_s} / N_0 N_n^* \\ &= \textstyle\sum_0^{\neq} E_{\text{lab}} \tilde{Y}_{[1]}^{a_1} \tilde{Y}_{[2]}^{a_2} \cdots \tilde{Y}_{[s]}^{a_s} / N_0 . \end{aligned}$$

By Definition 3.2, a term $\tilde{Y}_{[1]}^{a_1} \tilde{Y}_{[2]}^{a_2} \cdots \tilde{Y}_{[s]}^{a_s}$ would have to arise by relabelling the nested subscripts of the corresponding term $Y_{[1]}^{a_1} Y_{[2]}^{a_2} \cdots Y_{[s]}^{a_s}$ and since each value of a given nested subscript has an equal chance of being relabelled as a possible value of a crossed subscript within each combination of levels of the nesting subscripts, we have

$$E_{\text{lab}} \tilde{Y}_{[1]}^{a_1} \tilde{Y}_{[2]}^{a_2} \cdots \tilde{Y}_{[s]}^{a_s} = N_n^{-1} \textstyle\sum_n^{\neq} Y_{[1]}^{a_1} Y_{[2]}^{a_2} \cdots Y_{[s]}^{a_s} ,$$

where $\sum_n^{\neq}$ and $N_n$ denote the sum over the nested subscripts of $g(Y)$ and the number of terms in this sum respectively. Thus

$$\begin{aligned} E_{\text{lab}}g(\tilde{Y}) &= \textstyle\sum_0^{\neq} \sum_n^{\neq} Y_{[1]}^{a_1} Y_{[2]}^{a_2} \cdots Y_{[s]}^{a_s} / N_0 N_n \\ &= g(Y), \qquad \text{Q.E.D.} \end{aligned}$$

As a very simple example of this expectation over random cross labelling, let us consider the case where one factor, $B$ say, is nested in another factor $A$. Denote an observation from this population by $Y_{IJ}$ and further assume the range of each subscript is two units. The four observations might be arranged in a table as follows:

| $I=1$ | $I=2$ |
|---|---|
| $Y_{11}$ | $Y_{23}$ |
| $Y_{12}$ | $Y_{24}$ |

Now we randomly cross label the levels of $J$ within each level of $I$ thus forming the following possible crossed structures:

| | $I=1$ | $I=2$ |
|---|---|---|
| $J = 1$ | $\tilde{Y}_{11}(Y_{11})$ | $\tilde{Y}_{21}(Y_{23})$ |
| $J = 2$ | $\tilde{Y}_{12}(Y_{12})$ | $\tilde{Y}_{22}(Y_{24})$ |

| | $I=1$ | $I=2$ |
|---|---|---|
| $J = 1$ | $\tilde{Y}_{11}(Y_{11})$ | $\tilde{Y}_{21}(Y_{24})$ |
| $J = 2$ | $\tilde{Y}_{12}(Y_{12})$ | $\tilde{Y}_{22}(Y_{23})$ |

| | $I=1$ | $I=2$ |
|---|---|---|
| $J = 1$ | $\tilde{Y}_{11}(Y_{12})$ | $\tilde{Y}_{21}(Y_{23})$ |
| $J = 2$ | $\tilde{Y}_{12}(Y_{11})$ | $\tilde{Y}_{22}(Y_{24})$ |

| | $I=1$ | $I=2$ |
|---|---|---|
| $J = 1$ | $\tilde{Y}_{11}(Y_{12})$ | $\tilde{Y}_{21}(Y_{24})$ |
| $J = 2$ | $\tilde{Y}_{12}(Y_{11})$ | $\tilde{Y}_{22}(Y_{23})$, |

where the observations in parentheses denote the observations which were randomly labelled $\tilde{Y}_{IJ}$. If we take a gsm, say

$$\textstyle\sum^{\neq} \tilde{Y}_{IJ} \tilde{Y}_{I'J} / AB(A - 1) = \frac{1}{4}(\tilde{Y}_{11}\tilde{Y}_{21} + \tilde{Y}_{12}\tilde{Y}_{22} + \tilde{Y}_{21}\tilde{Y}_{11} + \tilde{Y}_{22}\tilde{Y}_{12}),$$

and average this over all possible crossed structures we obtain

$$\begin{aligned}
E_{\text{lab}}\tilde{Y}_{11}\tilde{Y}_{21} &= \tfrac{1}{4}(Y_{11}Y_{23} + Y_{11}Y_{24} + Y_{12}Y_{23} + Y_{12}Y_{24}),\\
E_{\text{lab}}\tilde{Y}_{12}\tilde{Y}_{22} &= \tfrac{1}{4}(Y_{12}Y_{24} + Y_{12}Y_{23} + Y_{11}Y_{24} + Y_{11}Y_{23}),\\
E_{\text{lab}}\tilde{Y}_{23}\tilde{Y}_{11} &= \tfrac{1}{4}(Y_{23}Y_{11} + Y_{24}Y_{11} + Y_{23}Y_{12} + Y_{24}Y_{12}),\\
E_{\text{lab}}\tilde{Y}_{22}\tilde{Y}_{12} &= \tfrac{1}{4}(Y_{24}Y_{12} + Y_{23}Y_{12} + Y_{24}Y_{11} + Y_{23}Y_{11}),
\end{aligned}$$

and thus

$$\begin{aligned}
E_{\text{lab}}[\textstyle\sum^{\neq} \tilde{Y}_{IJ}\tilde{Y}_{I'J}/AB(A-1)] &= \tfrac{1}{8}(Y_{11}Y_{23} + Y_{23}Y_{11} + Y_{11}Y_{24} + Y_{24}Y_{11}\\
&\qquad + Y_{12}Y_{23} + Y_{23}Y_{12} + Y_{12}Y_{24} + Y_{24}Y_{12})\\
&= \textstyle\sum^{\neq} Y_{IJ}Y_{I'J'}/AB^2(A-1).
\end{aligned}$$

Similarly we would find that

$$E_{\text{lab}}[\textstyle\sum^{\neq} \tilde{Y}_{IJ}\tilde{Y}_{I'J'}/AB(A-1)(B-1) = \textstyle\sum^{\neq} Y_{IJ}Y_{I'J'}/AB^2(A-1).$$

**6. Generalized polykays.** Following Hooke [2], we define a non-commutative dot-multiplication of gsm's for an $n$-way crossed structure.

DEFINITION 6.1. Let $\gamma_i$, $i = 1, \cdots, n$, denote arbitrary partitions of the set of symbols $p, q, r, \cdots$. Then

$$\begin{aligned}
\langle\gamma_1\rangle\cdot\langle\gamma_2\rangle\cdot \cdots \cdot\langle\gamma_n\rangle &= \langle\gamma_1/\gamma_2/\cdots/\gamma_n\rangle, && \text{if the } \gamma_i,\ i = 1, \cdots, n, \text{ consist of the same set of letters,}\\
&= 0, && \text{otherwise,}
\end{aligned}$$

where the $\langle\gamma_i\rangle$ represent simple symmetric means with respect to the individual factors.

This dot-multiplication is extended by distributivity to provide dot-multiplication for linear combinations of symmetric means.

DEFINITION 6.2. For an $n$-way crossed structure the generalized polykay $(\gamma_1/\gamma_2/\cdots/\gamma_n)$ is defined as

$$(\gamma_1/\gamma_2/\cdots/\gamma_n) = (\gamma_1)\cdot(\gamma_2)\cdot \cdots \cdot(\gamma_n)$$

where the $(\gamma_i)$, the simple polykays corresponding to the symmetric means $\langle\gamma_i\rangle$, are first expressed as linear sums of symmetric means before the dot-multiplication is taken.

DEFINITION 6.3. For an arbitrary population $P$ the possible generalized polykays are defined by reference to the crossed population $\tilde{P}$ obtained by random cross labelling of the nested subscripts of an observation in the following manner:

(i) take a polykay of the $n$-way structure and express it as a linear function of $n$-way gsm's;

(ii) take the expectation of the gsm's over all possible random cross labellings;

(iii) taking this expectation results in a linear function of gsm's of the arbitrary structure and if the expectation is non-zero, the $n$-way polykay is renamed, receiving the name of the leading gsm of the new polykay, and this result is the definition of that polykay for the structure $P$. If the expectation is zero, the $n$-way polykay of $\tilde{P}$ has no corresponding polykay in $P$.

Since the expectation of a gsm of $\tilde{P}$ is equal to its corresponding gsm of $P$, the polykay of $P$ corresponding to a polykay of $\tilde{P}$ may be formed simply by replacing each gsm of $\tilde{P}$ by its corresponding gsm's of $P$.

Because several gsm's of the $n$-way polykay may have the same expectation over random cross labelling, the $n$-way polykay is said to *collapse* under the expectation. Indeed, some of the polykays of $\tilde{P}$ vanish under this process.

**7. Sample symmetric means and polykays and inheritance on the average.** Thus far the gsm's and polykays have been defined with reference to population structures. Because of the one-to-one correspondence between the population and sample structures in pure sampling situations, sample gsm's can be defined exactly the same as in Definitions 3.2 and (3.2), replacing population values by sample values. The one-to-one correspondence between the population and sample structures is further elucidated by the following theorem:

THEOREM 7.1. *Let $f(Y)$ represent a given population* gsm *from an arbitrary structure $P$. Let $f(y)$ denote the corresponding sample* gsm. *Then*

$$E_{sa}f(y) = f(Y)$$

*where $E_{sa}$ refers to the expectation over the sampling procedure.*

PROOF. Suppose $f(y)$ is of the form

$$f(y) = d^{-1}\sum\nolimits^{\neq} y_{[1]}^{a_1} y_{[2]}^{a_2} \cdots y_{[s]}^{a_s} .$$

Further suppose $k$ factors are involved, denoted by $N_1, N_2, \cdots, N_k$, with sample ranges of $n_1, n_2, \cdots, n_k$ respectively. Let $N_1, N_2, \cdots, N_k$ also denote the population ranges of the corresponding population subscripts respectively. Consider the factor $N_i$. The $y$'s of $f(y)$ can be divided into several groups according to the likeness of priming of the subscripts corresponding to factors nesting $N_i$. Let the number of such groups be $r_i$. In this context, if the factor $N_i$ is not nested, we let $r_i = 1$. Further assume within each of the $r_i$ groups that there are $q_j^{i}$, $j = 1, \cdots, r_i$, differently primed subscripts corresponding to the factor $N_i$. Thus the denominator of $f(y)$ is

$$\begin{aligned} d &= \textstyle\prod_{i=1}^{k} [n_i(n_i - 1) \cdots (n_i - q_1^{i} + 1)][n_i(n_i - 1) \cdots (n_i - q_2^{i} + 1)] \\ &\qquad \cdots [n_i(n_i - 1) \cdots (n_i - q_{r_i} + 1)] \\ &= \textstyle\prod_{i=1}^{k} [n_i!/(n - q_1^{i})!] \cdots [n_i!/(n - q_{r_i}^{i} + 1)!]. \end{aligned}$$

The total number of possible samples from the population is $N = \prod_{i=1}^{k} \binom{N_i}{n_i}^{r_i}$ and each term $Y_{[1]}^{a_1} Y_{[2]}^{a_2} \cdots Y_{[s]}^{a_s}$ occurs in the same number of samples, this number being

$$N' = \prod\nolimits_{i=1}^{k} \binom{N_i - q_1^{i}}{n_i - q_1^{i}} \binom{N_i - q_2^{i}}{n_i - q_2^{i}} \cdots \binom{N_i - q_{r_i}^{i}}{n_i - q_{r_i}^{i}} .$$

Hence $E_{sa}f(y) = (N'/N\,d)\sum^{\neq} Y^{a_1}_{[1]}Y^{a_2}_{[2]} \cdots Y^{a_s}_{[s]}$ where the sum is, of course, over the population range of the subscripts. But

$$\begin{aligned}N'/N &= \textstyle\prod_{i=1}^{k} [(N_i - q_1{}^i)!/(N_i - n_i)!(n_i - q_1{}^i)!] \\ &\qquad \cdots [(N_i - q^i_{r_i})!/(N_i - n_i)!(N_i - q^i_{r_i})!]\cdot\{[(N_i - n_i)!]^{ri}(N_i\,!)^{ri}/(N_i\,!)^{ri}\} \\ &= \textstyle\prod_{i=1}^{k} [(N_i - q_1{}^i)!N_i!/N_i\,!(N_i - q_1{}^i)!] \\ &\qquad \cdots [(N_i - q^i_{r_i})!N_i\,!/N_i\,!(N_i - q^i_{r_i})!] \\ &= \frac{\prod_{i=1}^{k} [n_i(N_i - 1) \cdots (n_i q_1{}^i + 1)][n_i(n_i - 1) \cdots (n_i - q_1{}^i + 1)] \cdots [n_i(n_i - 1) \cdots (n_i - q^i_{r_i} + 1)]}{\prod_{i=1}^{k} [N_i(N_i - 1) \cdots (N_i - q_1{}^i + 1)][N_i(N_i - 1) \cdots (N_i - q_2{}^i + 1)] \cdots [N_i(N_i - 1) \cdots (N_i - q^i_{r_i} + 1)]},\end{aligned}$$

so that finally

$$\begin{aligned}&E_{sa}f(y) \\ &\quad = \frac{\sum^{\neq} Y^{a_1}_{[1]}\,Y^{a_2}_{[2]} \cdots Y^{a_s}_{[s]}}{\prod_{i=1}^{k} [N_i(N_i - 1) \cdots (N_i - q_1{}^i + 1)][N_i(N_i - 1) \cdots (N_i - q_2{}^i + 2)] \cdots [N_i(N_i - 1) \cdots (N_i - q^i_{r_i} + 1)]} \\ &\quad = f(Y).\end{aligned}$$

The sample polykays for the crossed structure are defined exactly as in Definition 6.2 again by replacing population values by sample values. In order to give the sample polykays from an arbitrary structure meaning of their own without reference to the population polykays we envisage the following sampling scheme:

(i) The arbitrary population $P$ is represented as a completely crossed structure by random cross labelling the levels of the nested factors.

(ii) One of the "artificially" induced crossed populations, say $\tilde{P}$, is selected randomly.

(iii) A crossed sample is now taken from $\tilde{P}$ obtaining a sample which would have been obtained by sampling directly from $P$.

To obtain the expectation of a function of the observations of a sample taken in the manner prescribed above, say $h(y)$, the following conditional expectation is required:

$$[Eh(y)] = E_{\text{lab}}\,E_{sa}[h(y)/\tilde{P}] \tag{7.1}$$

where $E_{sa}[h(y)/\tilde{P}]$ denotes the expectation of $h(y)$ over the possible ways of sampling a crossed sample for a fixed choice of a randomly cross labelled population, and $E_{\text{lab}}$ denotes the expectation over random cross labelling as described earlier.

It is obvious that if a sample is taken directly from the structure $P$ the subscripts of the sample observations can be random cross labelled and an expectation over random cross labelling of sample subscripts defined. Let $[h(y)/P']$

denote a sample gsm from a sample structure which has been randomly cross labelled, then it is easily seen that

$$E_{\text{lab}}E_{sa}[h(y)/\tilde{P}] = E_{sa}E_{\text{lab}}[h(y)/P'].$$

The sample polykays from an arbitrary structure are then defined exactly as in Definition 6.3 replacing population values by sample values.

Corollary 7.1. *Let $F(Y)$ represent a given population polykay from an arbitrary structure $P$. Let $F(y)$ denote the corresponding sample polykays. Then $E[F(y)] = F(Y)$.*

**8. Multiplication of gsm's and polykays.** The following theorem and corollary indicate a manner in which multiplication of gsm's and generalized polykays from a completely crossed structure can be effected, and form the basis for taking moments of generalized polykays from arbitrary structures.

Theorem 8.1. *Consider an n-way crossed sample structure. Denote two arbitrary sample* gsm's *by*

$$\langle\gamma_1/\gamma_2/\cdots/\gamma_n\rangle \quad \textit{and} \quad \langle\delta_1/\delta_2/\cdots/\delta_n\rangle.$$

*Then*

$$\langle\gamma_1/\gamma_2/\cdots/\gamma_n\rangle\langle\delta_1/\delta_2/\cdots/\delta_n\rangle = \langle\gamma_1\rangle\langle\delta_1\rangle\cdot\langle\gamma_2\rangle\langle\delta_2\rangle\cdot\ \cdots\ \cdot\langle\gamma_n\rangle\langle\delta_n\rangle$$

*where the $\langle\gamma_i\rangle\langle\delta_i\rangle$ represent the product of two partitions of the* ith *factor.*

Tukey [5] gives a scheme for obtaining these products.

Proof. Suppose $\langle\gamma_1/\gamma_2/\cdots/\gamma_n\rangle$ is of the form

$$(A_1A_2\cdots A_n)^{-1}\{\textstyle\sum^{\neq}_{\text{all}\,i,j,\cdots,p} y^{a_1}_{i\theta_1^1 j\theta_2^1\cdots p\theta_n^1}\, y^{a_2}_{i\theta_1^2 j\theta_2^2\cdots p\theta_n^2}\cdots y^{a_s}_{i\theta_1^s j\theta_2^s\cdots p\theta_n^s}\}$$

and $\langle\delta_1/\delta_2/\cdots/\delta_n\rangle$ is of the form

$$(B_1B_2\cdots B_n)^{-1}\{\textstyle\sum^{\neq}_{\text{all}\,l,m,\cdots,r} y^{b_1}_{l\alpha_1^1 m\alpha_2^1\cdots r\alpha_n^1}\, y^{b_2}_{l\alpha_1^2 m\alpha_2^2\cdots r\alpha_n^2}\cdots y^{b_s}_{l\alpha_1^s m\alpha_2^s\cdots p\alpha_n^s}\}$$

where $\sum_{i=1}^s a_i$ and $\sum_{i=1}^s b_i$ are not necessarily the same, and the $\alpha_v{}^u$ have the same meaning with respect to the subscripts $l, m, \cdots, r$ as the $\theta_v{}^u$ with respect to the subscripts $i, j, \cdots, p$ and are not necessarily different from the latter. The $A_i$ and $B_i$, $i = 1, \cdots, n$, denote the appropriate denominators for the first factor, second factor, etc. Then

$$\langle\gamma_1/\gamma_2/\cdots/\gamma_n\rangle\langle\delta_1/\delta_2/\cdots/\delta_n\rangle = \{\textstyle\sum^{\neq}_{\text{all}\,i,l}/A_1B_1\}\{\textstyle\sum^{\neq}_{\text{all}\,j,m}/A_2B_2)\}$$
$$\cdots\{\textstyle\sum^{\neq}_{\text{all}\,p,r} y^{a_1}_{i\theta_1^1 j\theta_2^1\cdots p\theta_n^1}\cdots y^{a_s}_{i\theta_1^s j\theta_2^s\cdots p\theta_n^s}\, y^{b_1}_{l\alpha_1^1 m\alpha_2^1\cdots r\alpha_n^1}\cdots y^{b_s}_{l\alpha_1^s m\alpha_2^s\cdots r\alpha_n^s}\}/A_nB_n\}.$$

But the last expression enclosed in braces, where the sum is over all the values of $p$ and $r$ with the restriction that subscripts differently primed must remain unequal, disregarding all subscripts but $p$ and $r$, represents the product $\langle\gamma_n\rangle\langle\delta_n\rangle$. This product results in some linear function of gsm's with respect to the $n$th factor. The partitioning with respect to the remaining factors is unchanged. Performing the second sum and division by $A_{n-1}B_{n-1}$ over each of the gsm's with respect to the $n$th factor is, by distribution of dot-multiplication, equivalent to the operation $\langle\gamma_{n-1}\rangle\langle\delta_{n-1}\rangle\cdot\langle\gamma_n\rangle\langle\delta_n\rangle$. Performing the next process of averaging over each of the resulting terms, now gsm's with respect to the $n$th and $(n - 1)$th factors, is equivalent to the operation $\langle\gamma_{n-2}\rangle\langle\delta_{n-2}\rangle\cdot\langle\gamma_{n-1}\rangle\langle\delta_{n-1}\rangle\cdot\langle\gamma_n\rangle\langle\delta_n\rangle$, etc. After the last averaging process, the complete operation $\langle\gamma_i\rangle\langle\delta_i\rangle\cdot\langle\gamma_2\rangle\langle\delta_2\rangle\cdot\ \cdots\ \cdot\langle\gamma_n\rangle\langle\delta_n\rangle$ will have been performed.

COROLLARY 8.1.

$$(\gamma_1/\gamma_2/\cdots/\gamma_n)(\delta_1/\delta_2/\cdots/\delta_n) = (\gamma_1)(\delta_1)\cdot(\gamma_2)(\delta_2)\cdot\ \cdots\ \cdot(\gamma_n)(\delta_n)$$

*where the* $(\gamma_i)(\delta_i)$ *represent products of simple sample polykays with respect to the* $i$th *factor.*

PROOF. This result obtains since each simple polykay can be expressed in terms of symmetric means and Theorem 7.3 can be applied to the products of symmetric means.

Although no general formulation will be made here it should be pointed out that each gsm from an arbitrary structure when randomly cross labelled can be expressed as a linear sum of gsm's from the completely crossed structure. The same is true of polykays. The latter is closely related to the fact that in the analysis of variance certain sums of squares in a nested analysis consist of the "pooling" of sums of squares from a crossed analysis. Thus to perform multiplication of gsm's from an arbitrary structure, one must first express these functions as linear sums of gsm's or polykays from a completely crossed structure, under random cross labelling, then perform the multiplication as indicated in the preceding theorem and corollary and then take the expectation of the result over the cross labelling.

**9. Some general comments.** Though a general formulation of generalized polykays has been presented in this paper, the application to specific examples may be rather limited. With the use of an electronic computer the author has generated all the distinct generalized polykays of degrees two, three and four for the two-factor crossed structure, the two-factor nested structure and five balanced three-factor structures. In addition, the author has obtained the variances and covariances of estimates of components of variation in these situations, the results in many cases being rather unwieldy. These results, as they are, will not be readily utilized until programs can be developed for computing values of the polykays in actual numerical examples—some research in this area is presently being done. The following section indicates the general approach used in obtaining the variances and covariances of the estimates of components of variation and illustrates the procedure in the case of the two-factor nested structure. The results for the two-factor crossed situation were given previously by Hooke [3].

**10. Variances and covariances of estimates of components of variation.** The $\Sigma$'s, defined by Zyskind [7] are defined in terms of the components of variation and thus the components of variation are expressible in terms of the $\Sigma$'s, which are in fact generalized polykays of degree two. The estimates of the components of variation are obtained by replacing the population polykays by the sample polykays. The problem of obtaining variances and covariances of these estimates is, then, that of obtaining variances and covariances of sample polykays. These results are functions of generalized polykays of degree four.

Denote by $F(y)$ a sample polykay of degree two from a completely crossed structure. Then the variance of $F(y)$ is

$$E\{F(y)^2 - E[F(y)]^2\} = E[F(y)]^2 - [F(Y)]^2.$$

Because of the property of inheritance on the average, $[F(Y)]^2$ can be obtained from $[F(y)]^2$ simply by replacing the sample polykays of degree four by the corresponding population polykays and replacing all sample values appearing as coefficients of the fourth degree polykays by corresponding population values.

A generalized sample polykay of degree two from an arbitrary structure $P$, say $H(y)$, can be expressed as a linear sum of polykays of degree two from a crossed structure. Assume $H(y)$ is of the form $H(y) = \sum_i [F_i(y) \mid P']$, where $P'$ represents the sample structure obtained by random cross labelling the sample subscripts of $P$ and $[F_i(y) \mid P']$ represents a second degree sample polykay from the crossed structure $P'$. Then

$$\begin{aligned} E[h(y)]^2 &= E\sum_i [F_i(y) \mid P']^2 \\ &= \sum_i E[F_i(y) \mid P']^2 + \sum_{i \neq i'} E[F_i(y) \mid P']\cdot[F_{i'}(y) \mid P'] \\ &= \sum_i E_{sa}E_{\text{lab}}[F_i(y) \mid P'] + \sum_{i \neq i'} E_{sa}E_{\text{lab}}[F_i(y) \mid P']\cdot[F_{i'}(y) \mid P'] \end{aligned}$$

from Equation 7.1. But the $[F_i(y) \mid P']^2$ and $[F_i(y) \mid P']\cdot[F_{i'}(y) \mid P']$ are obtained by the multiplication formula for crossed polykays and hence the desired is obtained by taking the expectations of these products over random cross labelling and sampling.

As an example, consider the model case where a balanced sample is taken from a two-factor nested structure. The analysis of variance table may be represented as follows:

| Source | d.f. | E.M.S. |
|---|---|---|
| $A$ | $a - 1$ | $\Sigma_{A(B)} + b\Sigma_A$ |
| $A(B)$ | $a(b - 1)$ | $\Sigma_{A(B)}$ |

where $\Sigma_{A(B)} = \sigma^2_{AB}$, $\Sigma_A = \sigma_A{}^2 - B^{-1}\sigma^2_{A(B)}$, it being understood that the $\Sigma$'s are in fact second degree generalized polykays. Then

$$\sigma^2_{A(B)} = \Sigma_{A(B)}, \qquad \sigma_A{}^2 = \Sigma_A + B^{-1}\Sigma_{A(B)},$$

and

$$\begin{aligned} \text{Var}\ (\hat{\sigma}^2_{A(B)}) &= \text{Var}\ (\Sigma_{a(b)}), \\ \text{Var}\ (\hat{\sigma}_A{}^2) &= \text{Var}\ (\Sigma_a) + B^{-1}\ \text{Cov}\ (\Sigma_a, \Sigma_{a(b)}) + B^{-2}\ \text{Var}\ (\Sigma_{a(b)}), \\ \text{Cov}\ (\hat{\sigma}_A{}^2, \hat{\sigma}^2_{A(B)}) &= \text{Cov}\ (\Sigma_a, \Sigma_{a(b)}) + B^{-1}\ \text{Var}\ (\Sigma_{a(b)}). \end{aligned}$$

To obtain the variances and covariances of the Σ's in this situation we first note that under random cross labelling we can write $\Sigma_a = \Sigma_a'$, $\Sigma_{a(b)} = \Sigma_b' + \Sigma_{ab}'$, where the primes denote Σ's for the two-factor crossed structure

$$(\Sigma_A' = \sigma_A^2 - B^{-1}\sigma_{AB}^2 ; \qquad \Sigma_B' = \sigma_B^2 - A^{-1}\sigma_{AB}^2 ; \qquad \Sigma_{AB}' = \sigma_{AB}^2).$$

This result is easily verified by observing that the sum of squares for $A(B)$ is actually a pooled sum of squares (under cross labelling of subscripts) i.e., S.S. $A(B)$ = S.S. $B'$ + S.S. $AB'$. Representing the sums of squares by sample Σ's leads to the desired result.

Now the variances of these sample Σ's from the crossed structure (Hooke's bipolykays) have been obtained previously and will not be illustrated here. The variances and covariances of the Σ's from the nested structure are then obtained by taking the expectation of the variances of the Σ's from the cross structure over random cross labelling. This, of course, is equivalent to replacing the fourth degree polykays of the crossed structure by the corresponding polykays of the nested structure. Thus

$$\begin{aligned}
\text{Var}\ (\Sigma_{a(b)}) &= E_{\text{lab}}[\text{Var}\ (\Sigma_b') + \text{Var}\ (\Sigma_{ab}') + 2\ \text{Cov}\ (\Sigma_b', \Sigma_{ab}')],\\
\text{Var}\ (\Sigma_a) &= E_{\text{lab}}(\Sigma_{\iota}'),\\
\text{Cov}\ (\Sigma_a , \Sigma_{a(b)}) &= E_{\text{lab}}[\text{Cov}\ (\Sigma_a', \Sigma_b') + \text{Cov}\ (\Sigma_a', \Sigma_{ab}')].
\end{aligned}$$

Performing the indicated operation leads to the following:

$$\begin{aligned}
\text{Var}\ (\Sigma_{a(b)}) = {} & (2/a(b-1) - 2/A(B-1))(AABB/A_1A_1B_2B_2)\\
& + [(a^{-1} + A^{-1}) + (2/a(b-1) - 2/A(B-1))]\\
& \quad \cdot (AAAA/A_1A_1B_1B_1)\\
& + (1/ab - 1/AB)(AAAA/A_1A_1A_1A_1),\\
\text{Var}\ (\Sigma_a) = {} & (2/(a-1) - 2/(A-1))(AABB/A_1B_1C_2D_2)\\
& + (a^{-1} - A^{-1})(AAAA/A_1B_1C_1D_1)\\
& + (4/b(a-1) - 4/B(A-1))(AABB/A_1A_1B_2C_2)\\
& + (4/ab - 4/AB)(AAAA/A_1A_1B_1C_1)\\
& + (2/b(a-1)(b-1) - 2/B(A-1)(B-1))\\
& \quad \cdot (AABB/A_1A_1B_2B_2)\\
& + (2/ab(b-1) - 2/AB(B-1))(AAAA/A_1A_1B_1B_1)\\
\text{Cov}\ (\Sigma_a , \Sigma_{a(b)}) = {} & (a^{-1} - A^{-1}(AAAA/A_1A_1B_1C_1)\\
& - (2/ab(b-1) - 2/AB(B-1))(AABB/A_1A_1B_2B_2)\\
& - (2/ab(b-1) - 2/AB(B-1))(AAAA/A_1A_1B_1B_1)\\
& + (2/ab - 2/AB)(AAAA/A_1A_1A_1B_1),
\end{aligned}$$

where

$$\begin{aligned}
(AAAA/A_1A_1A_1A_1) &= \langle AAAA/A_1A_1A_1A_1\rangle - 4\langle AAAA/A_1A_1A_1B_1\rangle \\
&\quad - 3\langle AAAA/A_1A_1B_1B_1\rangle + 12\langle AAAA/A_1A_1B_1C_1\rangle \\
&\quad - 6\langle AAAA/A_1B_1C_1D_1\rangle, \\
(AAAA/A_1A_1A_1B_1) &= \langle AAAA/A_1A_1A_1B_1\rangle - \langle AAAB/A_1A_1A_1B_2\rangle \\
&\quad - 3\langle AAAA/A_1A_1B_1C_1\rangle + 3\langle AAAB/A_1A_1B_1C_2\rangle \\
&\quad + 2\langle AAAA/A_1B_1C_1D_1\rangle - 2\langle AAAB/A_1B_1C_1D_2\rangle, \\
(AAAA/A_1A_1B_1B_1) &= \langle AAAA/A_1A_1B_1B_1\rangle - \langle AABB/A_1A_1B_2B_2\rangle \\
&\quad - 2\langle AAAA/A_1A_1B_1C_1\rangle + 2\langle AABB/A_1A_1B_2C_2\rangle \\
&\quad + \langle AAAA/A_1B_1C_1D_1\rangle - \langle AABB/A_1B_1C_2D_2\rangle, \\
(AABB/A_1A_1B_2B_2) &= \langle AABB/A_1A_1B_2B_2\rangle - 2\langle AABB/A_1A_1B_2C_2\rangle \\
&\quad + \langle AABB/A_1B_1C_2D_2\rangle, \\
(AAAA/A_1A_1B_1C_1) &= \langle AAAA/A_1A_1B_1C_1\rangle - 2\langle AAAB/A_1A_1B_1C_2\rangle \\
&\quad - \langle AABB/A_1A_1B_2C_2\rangle + 2\langle AABC/A_1A_1B_2C_3\rangle \\
&\quad - \langle AAAA/A_1B_1C_1D_1\rangle + 2\langle AAAB/A_1B_1C_1D_2\rangle \\
&\quad + \langle AABB/A_1B_1C_2D_2\rangle - 2\langle AABC/A_1B_1C_2D_3\rangle, \\
(AABB/A_1A_1B_2C_2) &= \langle AABB/A_1A_1B_2C_3\rangle - \langle AABC/A_1A_1B_2C_3\rangle \\
&\quad - \langle AABB/A_1B_1C_2D_2\rangle + \langle AABC/A_1B_1C_2D_3\rangle, \\
(AAAA/A_1B_1C_1D_1) &= \langle AAAA/A_1B_1C_1D_1\rangle - 4\langle AAAB/A_1B_1C_1D_2\rangle \\
&\quad - 3\langle AABB/A_1B_1C_2D_2\rangle + 12\langle AABC/A_1B_1C_2D_3\rangle \\
&\quad - 6\langle ABCD/A_1B_2C_3D_4\rangle, \\
(AABB/A_1B_1C_2D_2) &= \langle AABB/A_1B_1C_2D_2\rangle - 2\langle AABC/A_1B_1C_2D_3\rangle \\
&\quad + \langle ABCD/A_1B_2C_3D_4\rangle.
\end{aligned}$$

**11. Acknowledgment.** The author wishes to acknowledge Professor Oscar Kempthorne of Iowa State University for his guidance during the research on this project and for his assistance in formulating the concept of random cross labelling introduced in this paper.

## REFERENCES

[1] Dayhoff, Eugene (1964). On the equivalence of polykays of the second degree and $\Sigma$'s. *Ann. Math. Statist.* **35** 1663–1672.

[2] Hooke, Robert (1956). Symmetric functions of a two-way array. *Ann. Math. Statist.* **27** 55–79.

[3] Hooke, Robert (1956). Some applications of bipolykays to the estimation of variance components and their moments. *Ann. Math. Statist.* **27** 80–98.

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

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

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

[7] Zyskind, George (1962). On structure, relation, Σ, and expectation of mean squares. *Sankyā Ser. A* **24** 115–147.