# MEASUREMENT OF DIVERSITY: MULTIPLE CELL CONTENTS

F. N. DAVID
UNIVERSITY OF CALIFORNIA, RIVERSIDE

## 1. Introduction

In a previous paper [1], we discussed the distribution of a score, representing a diversity index, when a number of different colored balls are randomly dropped into $M(=mn)$ identical compartments, one ball only permitted per compartment. We now generalize and at the same time simplify the previous setup by allowing more than one ball per box.

## 2. Notation

A box of $m \times n = M$ compartments is supposed. There will be $(m-1)(n-1)$ crossover points each of which will be surrounded by four compartments. Denote these crossover points by $(ij)$, $i = 1, 2, \cdots, m-1$, $j = 1, 2, \cdots, n-1$. Let there be $K_1$ balls in $s$ colors with $k_\ell$ the number of balls of the $\ell$th color and

$$\sum_{\ell=1}^{s} k_\ell = K_1. \tag{2.1}$$

These $K_1$ balls are supposedly dropped randomly into the $M$ compartments, with no limitation on the individual compartment capacity. Consider the $(ij)$th crossover point. Let $T_{ij}$ be the total number of balls in the four compartments surrounding $(ij)$. Let $t_{ij\ell}$ be the total number of balls of the $\ell$th color in the same four compartments so that

$$T_{ij} = \sum_{\ell=1}^{s} t_{ij\ell}. \tag{2.2}$$

The number of joins between balls of like and unlike colors will be, omitting the factor of one half,

$$T_{ij}^{(2)} = \sum_{\ell=1}^{s} t_{ij\ell}^{(2)} + \sum_{\ell \neq h} t_{ij\ell} t_{ijh}, \tag{2.3}$$

for the four boxes. Summed for all values of $i$ and $j$, we have

$$\sum_{i=1}^{m-1} \sum_{j=1}^{n-1} T_{ij}^{(2)} = \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \sum_{\ell=1}^{s} t_{ij\ell}^{(2)} + \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \sum_{\ell \neq h} t_{ij\ell} t_{ijh}. \tag{2.4}$$

This investigation was partially supported by USPHS Research Grant No. GM-10525-08 National Institutes of Health, Public Health Service.

Let us write this as $S_T = S_W + S_B$, a break up of the joins scores reminiscent of the classical analysis of variance setup. Clearly, $S_W$ represents a measure of the within color joins, while $S_B$ gives the between color joins. Accordingly, $S_W$ may be used as a measure of aggregation (or clumping), while $S_B$ may be used to measure segregation (or affinity). A modification of $S_B$ is suggested below.

## 3. Conditional means

The restriction that there is no limit on the individual compartment capacity means that, for any four boxes considered, the number in these four boxes of any one given color will be a positive binomial variable with index $k_\ell$, $\ell = 1, 2, \cdots, s$, and probability $p = 4/M$. The restriction is not very important and the procedure can be modified if desired by placing an upper content limit on each box. The supposition that the balls are randomly dropped implies an independence between and within color content for the four boxes, but because of the method of scoring $S_W$ and $S_B$ will be correlated. Conditional on the numbers $k_\ell$, $\ell = 1, 2, \cdots, s$, and $K_1$, we have

$$\begin{aligned} \mathcal{E}(S_T) &= (m-1)(n-1)K_1^{(2)}p^2, \\ \mathcal{E}(S_W) &= (m-1)(n-1)K_2p^2, \\ \mathcal{E}(S_B) &= (m-1)(n-1)[K_1^{(2)} - K_2]p^2, \end{aligned} \tag{3.1}$$

where, following our previous notation,

$$K_r = \sum_{\ell=1}^{s} k_\ell^{(r)}. \tag{3.2}$$

When $n = 1$, that is, when the boxes are in a line, the score will be taken for each of two connected boxes, $p$ will be $2/m$ and

$$\mathcal{E}(S_T) = (m-1)K_1^{(2)}p^2 \tag{3.3}$$

with the $S_W$ and $S_B$ similarly modified. It may also be noted that for three dimensions there will be a three dimensional lattice formed by the intersections of the box edges. The score will be calculated from the contents of the eight boxes surrounding the $(ijw)$ node, $p$ will be equal to $8/M$ where $M = m \times n \times w$, $w$ being the number of boxes in the third dimension. Accordingly, for three dimensions,

$$\mathcal{E}(S_T) = (m-1)(n-1)(w-1)K_1^{(2)}p^2 \tag{3.4}$$

with similar expressions for the other two sums.

## 4. Unconditional means

Biologically measures conditional on the $\{k_\ell\}$ seem to be those sought after. However the $k$ may themselves have arisen from a larger set and so the unconditional moments may be of interest. Suppose free multinomial sampling with

$$P\left\{\prod_{\ell=1}^{s} k_\ell\right\} = \frac{K_1!}{\prod_{\ell=1}^{s} k_\ell!} \prod_{\ell=1}^{s} p_\ell^{*k_\ell}, \tag{4.1}$$

and write

$$P_r = \sum_{\ell=1}^{s} p_\ell^{*r}, \qquad P_1 = 1. \tag{4.2}$$

Then we have

$$\begin{aligned} \mathcal{E}(S_W) &= (m-1)(n-1)K_1^{(2)}P_2p^2, \\ \mathcal{E}(S_B) &= (m-1)(n-1)p^2K_1^{(2)}(1-P_2), \end{aligned} \tag{4.3}$$

with

$$\mathcal{E}(S_T) = (m-1)(n-1)p^2K_1^{(2)} \tag{4.4}$$

as before. The modifications for one or three dimensions instead of the two dimensional result above are easily written down.

Instead of multinomial sampling, we may wish to suppose a multiple Pólya urn model. Let there be an urn containing $N$ balls of $s$ different colors with $R_\ell$ balls of the $\ell$th color, $\ell = 1, 2, \cdots, s$. A ball is drawn, the color noted, and it is returned to the urn together with $\Delta$ balls of the same color. Writing $\delta = \Delta/N$, $p_\ell^* = R_\ell/N$, we have

$$P(\{k_\ell\}) = \binom{K_1}{k_1 \cdots k_s} \frac{\prod_{\ell=1}^{s}\prod_{h=1}^{k_\ell} (p_\ell^* + (h-1)\delta)}{\prod_{a=1}^{K_1} (1 + (a-1)\delta)}. \tag{4.5}$$

If the sampling is of this type and if

$$P_r = \sum_{\ell=1}^{s} \frac{p_\ell^*(p_\ell^* + \delta) \cdots (p_\ell^* + (r-1)\delta)}{1(1+\delta) \cdots (1 + (r-1)\delta)}, \tag{4.6}$$

then the expectations are formally as above.

## 5. Conditional variances (two dimensions)

As in our previous paper it proved advantageous to evaluate the second crude moment and to split the evaluation into four different steps. Thus, we have

$$S_W^2 = \sum_i \sum_j \sum_r \sum_s \left\{\sum_\ell t_{ij\ell}^{(2)} t_{rs\ell}^{(2)} + \sum_{\ell \neq h} t_{ij\ell}^{(2)} t_{rsh}^{(2)}\right\}. \tag{5.1}$$

*Case (i): $i = r, j = s$; $(m-1)(n-1)$ terms.*

$$\sum_\ell (t_{ij\ell}^{(2)})^2 = \sum_\ell [t_{ij\ell}^{(4)} + 4t_{ij\ell}^{(3)} + 2t_{ij\ell}^{(2)}], \tag{5.2}$$

which has expectation

$$K_4p^4 + 4K_3p^3 + 2K_2p^2. \tag{5.3}$$

Again,

$$\sum_{\ell \neq h} t_{ij\ell}^{(2)} t_{ijh}^{(2)} = p^4 \sum_{\ell \neq h} k_\ell^{(2)} k_h^{(2)} = p^4\{K_2^2 - K_4 - 4K_3 - 2K_2\}, \tag{5.4}$$

so that the total contribution is

$$p^4[K_2^2 - 4K_3 - 2K_2] + 4p^3K_3 + 2p^2K_2. \tag{5.5}$$

*Case (ii):* $r = i + 1$, $j = s$ or $i = r$, $s = j + 1$; $2(m - 2)(n - 1)$ *or* $2(m - 1)(n - 2)$ *terms.* The total contribution is

$$p^4[K_2^2 - 4K_3 - 2K_2] + 2p^3K_3 + \tfrac{1}{2}p^2K_2. \tag{5.6}$$

*Case (iii):* $r = i + 1$, $s = j + 1$; $4(m - 2)(n - 2)$ *terms.* The total contribution is

$$p^4[K_2^2 - 4K_3 - 2K_2] + p^3K_3 + \tfrac{1}{8}p^2K_2. \tag{5.7}$$

*Case (iv):* $|i - r| \geqq 2$, $|j - s| \geqq 2$; $[(m - 1)^2(n - 1)^2 - (3m - 5)(3n - 5)]$ *terms.* The total contribution is

$$p^4[K_2^2 - 4K_3 - 2K_2]. \tag{5.8}$$

Collecting terms, we have

$$\begin{aligned} \mathcal{E}(S_W^2) = {} & (m - 1)^2(n - 1)^2p^4[K_2^2 - 4K_3 - 2K_2] \\ & + 4p^3K_3(2m - 3)(2n - 3) + \tfrac{1}{2}p^2K_2(3m - 4)(3n - 4) \end{aligned} \tag{5.9}$$

with

$$\begin{aligned} \operatorname{Var} S_W = {} & -(m - 1)^2(n - 1)^2p^4(4K_3 + 2K_2) \\ & + 4p^3K_3(2m - 3)(2n - 3) + \tfrac{1}{2}p^2K_2(3m - 4)(3n - 4). \end{aligned} \tag{5.10}$$

Similarly,

$$\begin{aligned} \operatorname{Var} S_T = {} & -(m - 1)^2(n - 1)^2p^4(4K_1^{(3)} + 2K_1^{(2)}) \\ & + 4p^3K_1^{(3)}(2m - 3)(2n - 3) + \tfrac{1}{2}p^2K_1^{(2)}(3m - 4)(3n - 4), \end{aligned} \tag{5.11}$$

$$\begin{aligned} & \operatorname{Var} S_B \\ & = -(m - 1)^2(n - 1)^2p^4[4K_1^{(3)} + 2K_1^{(2)} + (4K_3 + 2K_2) - 4K_2(2K_1 - 3)] \\ & \quad + 4(2m - 3)(2n - 3)p^3[K_1^{(3)} + K_3 - 2K_2(K_1 - 2)] \\ & \quad + \tfrac{1}{2}p^2(3m - 4)(3n - 4)[K_1^{(2)} - K_2], \end{aligned} \tag{5.12}$$

and

$$\begin{aligned} & \operatorname{Cov} S_WS_B \\ & = 4[K_3 - K_2(K_1 - 2)][p^4(m - 1)^2(n - 1)^2 - p^3(2m - 3)(2n - 3)]. \end{aligned} \tag{5.13}$$

## 6. Conditional variances (one and three dimensions)

Remembering that $p = 2/m$ for one dimension,

$$\operatorname{Var} S_W = -(m - 1)^2p^4(4K_3 + 2K_2) + 4p^3K_3(2m - 3) + p^2K_2(3m - 4) \tag{6.1}$$

with similar modifications for the others, namely,

$$\begin{aligned} \operatorname{Var} S_B = {} & -(m - 1)^2p^4(4K_1^{(3)} + 4K_3 - 8K_2K_1 + 14K_2 + 2K_1^{(2)}) \\ & + 4p^3(2m - 3)[K_1^{(3)} + K_3 - 2K_2K_1 + 4K_2] \\ & + p^2(3m - 4)[K_1^{(2)} - K_2], \end{aligned} \tag{6.2}$$

$$\operatorname{Cov}(S_WS_B) = 4[K_3 - K_2(K_1 - 2)][p^4(m - 1)^2 - p^3(2m - 3)]. \tag{6.3}$$

For three dimensions $m$, $n$ and $w$, we have that $p = 8/m\cdot n\cdot w$. The functions of $K$ remain as before with suitable modifications of the multiplying factors.

## 7. Unconditional variances (two dimensions)

The unconditional variances follow along lines similar to those for the means. With free multinomial sampling, we have

$$(7.1)\quad \mathcal{E}(S_W^2) = K_1^{(4)}P_2^2p^4(m-1)^2(n-1)^2 + 4K_1^{(3)}P_3p^3(2m-3)(2n-3) + \tfrac{1}{2}K_1^{(2)}P_2p^2(3m-4)(3n-4),$$

giving

$$(7.2)\quad \operatorname{Var} S_W = -(4K_1^{(3)} + 2K_1^{(2)})P_2^2p^4(m-1)^2(n-1)^2 + 4K_1^{(3)}P_3p^3(2m-3)(2n-3) + \tfrac{1}{2}K_1^{(2)}P_2p^2(3m-4)(3n-4).$$

Similarly,

$$(7.3)\quad \begin{aligned} \operatorname{Var} S_B &= -(4K_1^{(3)} + 2K_1^{(2)})(1-P_2)^2p^4(m-1)^2(n-1)^2 \\ &\quad + 4K_1^{(3)}p^3(2m-3)(2n-3)[(P_3 - 2P_2 + 1) \\ &\quad + \tfrac{1}{2}K_1^{(2)}p^2(3m-4)(3n-4)(1-P_2)], \\ \operatorname{Cov} S_WS_B &= -(4K_1^{(3)} + 2K_1^{(2)})P_2(1-P_2)p^4(m-1)^2(n-1)^2 \\ &\quad + 4K_1^{(3)}p^3(2m-3)(2n-3)[-P_3 + P_2]. \end{aligned}$$

The Pólya second order moments may be calculated from first principles but are possibly more simply achieved from the $k$-function technique of the next section.

## 8. Augmented factorial monomial symmetric procedure

As previously written, $k_\ell$ stands for the number of the $\ell$th kind to be distributed. The suffix is, however, not important and in cases where there is no possibility of confusion it may be omitted, following the usual symmetric function procedure. For the sake of example, let us suppose we are interested in the sum of the $r$th factorial powers of all the $k$. We may write this sum as follows:

$$(8.1)\quad \sum_{\ell=1}^{s} k_\ell^{(r)} = [k_r].$$

Again, we may write

$$(8.2)\quad \sum_{\ell=1}^{s}\sum_{h=1}^{s} k_\ell^{(r)}k_h^{(r)} = [k_r^2].$$

Accordingly for the first moments of $S_W$ and $S_B$, respectively, we may write

$$(8.3)\quad \begin{aligned} \mu_{10}' &= (m-1)(n-1)p^2K_2 = (m-1)(n-1)p^2[k_2], \\ \mu_{01}' &= (m-1)(n-1)p^2[K_1^{(2)} - K_2]p^2 = (m-1)(n-1)p^2[k_1^2], \end{aligned}$$

since

$$(8.4)\quad \begin{aligned} [k_1^2] &= \sum_{\ell\neq h} k_\ell k_h = (\sum k_\ell)^2 - \sum k_\ell^2 = K_1^2 - [\sum k_\ell^{(2)} + \sum k_\ell] \\ &= K_1^{(2)} - K_2. \end{aligned}$$

This notation enables complicated multiplications to be carried out quickly and lessens the risk of error. Thus, we have, for two dimensions,

$$(8.5)\quad \mu_{20} = p^4(m-1)^2(n-1)^2[-4[k_3] - 2[k_2]] + 4p^3(2m-3)(2n-3)[k_3] + \tfrac{1}{2}p^2(3m-4)(3n-4)[k_2],$$

$$(8.6)\quad \mu_{11} = 4[k_2k_1][-p^4(m-1)^2(n-1)^2 + p^3(2m-3)(2n-3)],$$

$$(8.7)\quad \mu_{02} = p^4(m-1)^2(n-1)^2[-4[k_2k_1] - 4[k_1^3] - 2[k_1^2]] + 4p^3(2m-3)(2n-3)[[k_2k_1] + [k_1^3]] + \tfrac{1}{2}p^2(3m-4)(3n-4)[k_1^2],$$

giving as a check

$$(8.8)\quad \operatorname{Var} S_T = -p^4(m-1)^2(n-1)^2\{4[[k_3] + 3[k_2k_1] + [k_1^3]] + 2[[k_2] + [k_1^2]]\} + 4p^3(2m-3)(2n-3)\{[k_3] + 3[k_2k_1] + [k_1^3]\} + \tfrac{1}{2}p^2(3m-4)(3n-4)\{[k_2] + [k_1^2]\},$$

or

$$(8.9)\quad \operatorname{Var} S_T = -p^4(m-1)^2(n-1)^2\{4K_1^{(3)} + 2K_1^{(2)}\} + 4p^3(2m-3)(2n-3)K_1^{(3)} + \tfrac{1}{2}p^2(3m-4)(3n-4)K_1^{(2)}$$

as before.

In a previous paper [1] when I introduced these $k$-functions I demonstrated how the augmented monomial symmetric function tables ([2] or [3]) could, formally, be used to evaluate the $k$-functions in terms of the $K$-functions, and gave expressions for the products of the $K$-functions corresponding to the products of the usual augmented monomial symmetrics. From these tables (see Table IV for illustrations), we read off the $K$-functions from the $k$ experssions and writing for brevity,

$$(8.10)\quad \begin{matrix} X = K_3, & Y = K_2(K_1 - 2), & Z = K_1^{(3)}, \\ x = K_2, & y = K_1^{(2)}; & \end{matrix}$$

we have, as already demonstrated,

$$(8.11)\quad \mu_{10}' = (m-1)(n-1)p^2\cdot x, \qquad \mu_{01}' = (m-1)(n-1)p^2[y - x],$$

$$(8.12)\quad \mu_{20} = -p^4(m-1)^2(n-1)^2[4X + 2x] + 4p^3(2m-3)(2n-3)X + \tfrac{1}{2}p^2(3m-4)(3n-4)x,$$

$$(8.13)\quad \mu_{11} = -4[-p^4(m-1)^2(n-1)^2 + p^3(2m-3)(2n-3)][X - Y],$$

$$(8.14)\quad \mu_{02} = -p^4(m-1)^2(n-1)^2[4(X - 2Y + Z) - 2(x - y)] + 4p^3(2m-3)(2n-3)[X - 2Y + Z] + \tfrac{1}{2}p^2(3m-4)(3n-4)[-(x - y)].$$

## 9. Conditional third moments (two dimensions)

Using the $k$-function technique and writing

$$(9.1)\quad \begin{matrix} \alpha = K_4, & \beta = K_3(K_1 - 3), & \gamma = K_2(K_2 - 2) - 4K_3, \\ & \delta = K_2(K_1 - 2)^{(2)}, & \varepsilon = K_1^{(4)} \end{matrix}$$

with $X$, $Y$, $Z$, $x$, and $y$ as defined above, we may construct Table I.

## TABLE I

### Conditional Third Moments

Notation:

$$\mu_{30} = \mathcal{E}(S_W - \mathcal{E}(S_W))^3,\ \mu_{21} = \mathcal{E}(S_W - \mathcal{E}(S_W))^2(S_B - \mathcal{E}(S_B)),\ \mu_{12} = \mathcal{E}(S_W - \mathcal{E}(S_W))(S_B - \mathcal{E}(S_B))^2,\ \mu_{03} = \mathcal{E}(S_B - \mathcal{E}(S_B))^3$$

$$\alpha = K_4,\ \beta = K_3(K_1 - 3),\ \gamma = K_2(K_2 - 2) - 4K_3,\ \delta = K_2(K_1 - 2)^{(2)},\ \varepsilon = K_1^{(4)}$$

$$X = K_3,\ Y = K_2(K_1 - 2),\ Z = K_1^{(3)}$$

$$x = K_2,\ y = K_1^{(2)}$$

| Central Moments | Multipliers of Tabular Entries $p^6(m-1)^3(n-1)^3$ | $p^5(m-1)(n-1)(2m-3)(2n-3)$ | $p^4(m-1)(n-1)(3m-4)(3n-4)$ |
|---|---|---|---|
| $\mu_{30}$ | $40\alpha + 64X + 8x$ | $-72\alpha - 72X$ | $-6X - 3x$ |
| $\mu_{21}$ | $-40\alpha + 32\beta + 8\gamma + 16(-X + Y)$ | $72\alpha - 56\beta - 16\gamma + 16(X + Y)$ | $2(X - Y)$ |
| $\mu_{12}$ | $40\alpha - 64\beta - 16\gamma + 40\delta - 32(X - Y)$ | $-72\alpha + 112\beta + 32\gamma - 72\delta + 40(X - Y)$ | $2(X - Y)$ |
| $\mu_{03}$ | $-40\alpha + 96\beta + 24\gamma - 120\delta + 40\varepsilon + 80X - 144Y + 64Z + 8(-x + y)$ | $72\alpha - 168\beta - 48\gamma + 216\delta - 72\varepsilon - 96X + 168Y - 72Z$ | $-6X + 12Y - 6Z + 3(x - y)$ |
| $\mu_{3T}$ | $40\varepsilon + 64Z + 8y$ | $-72\varepsilon - 72Z$ | $-6Z - 3y$ |

| Central Moments | Multipliers of Tabular Entries $p^4(8m-15)(8n-15)$ | $p^4(4m-7)(4n-7)$ | $p^3(3m-5)(3n-5)$ | $p^3(5m-8)(5n-8)$ | $p^2(5m-8)(5n-8)$ |
|---|---|---|---|---|---|
| $\mu_{30}$ | $6\alpha$ | $8\alpha$ | $24X$ | $2X$ | $x$ |
| $\mu_{21}$ | $-6\alpha + 4\beta + 2\gamma$ | $-8\alpha + 8\beta$ | $-8(X - Y)$ | — | — |
| $\mu_{12}$ | $6\alpha - 8\beta - 4\gamma + 6\delta$ | $8\alpha - 16\beta + 8\delta$ | $-8(X - Y)$ | $-2(X - Y)$ | — |
| $\mu_{03}$ | $-6\alpha + 12\beta + 6\gamma - 18\delta + 6\varepsilon$ | $-8\alpha + 24\beta - 24\delta + 8\varepsilon$ | $24X - 48Y - 24Z$ | $4X - 6Y + 2Z$ | $-(x - y)$ |
| $\mu_{3T}$ | $6\varepsilon$ | $8\varepsilon$ | $24Z$ | $2Z$ | $y$ |

TABLE II

NUMBERS OF TERMS IN THE EXPANSION OF EXPRESSIONS SUCH AS $\left[\sum_i \sum_j \alpha_{ij}\right]^3$

| | $i^3$ | $i^2(i+1)$ | $i^2r$ | $i(i+1)(i+2)$ | $i(i+1)r$ | $ihr$ | Total |
|---|---|---|---|---|---|---|---|
| $j^3$ | $(m-1)(n-1)$ | $6(m-2)(n-1)$ | $3(m-2)^{(2)}(n-1)$ | $6(m-3)(n-1)$ | $6(m-3)^{(2)}(n-1)$ | $(m-3)^{(3)}(n-1)$ | $(m-1)^3(n-1)$ |
| $j^2(j+1)$ | $6(m-1)(n-2)$ | $36(m-2)(n-2)$ | $18(m-2)^{(2)}(n-2)$ | $36(m-3)(n-2)$ | $36(m-3)^{(2)}(n-2)$ | $6(m-3)^{(3)}(n-2)$ | $6(m-1)^3(n-2)$ |
| $j^2s$ | $3(m-1)(n-2)^{(2)}$ | $18(m-2)(n-2)^{(2)}$ | $9(m-2)^{(2)}(n-2)^{(2)}$ | $18(m-3)(n-2)^{(2)}$ | $18(m-3)^{(2)}(n-2)^{(2)}$ | $3(m-3)^{(3)}(n-2)^{(2)}$ | $3(m-1)^3(n-2)^{(2)}$ |
| $j(j+1)(j+2)$ | $6(m-1)(n-3)$ | $36(m-2)(n-3)$ | $18(m-2)^{(2)}(n-3)$ | $36(m-3)(n-3)$ | $36(m-3)^{(2)}(n-3)$ | $6(m-3)^{(3)}(n-3)$ | $6(m-1)^3(n-3)$ |
| $j(j+1)s$ | $6(m-1)(n-3)^{(2)}$ | $36(m-2)(n-3)^{(2)}$ | $18(m-2)^{(2)}(n-3)^{(2)}$ | $36(m-3)(n-3)^{(2)}$ | $36(m-3)^{(2)}(n-3)^{(2)}$ | $6(m-3)^{(3)}(n-3)^{(2)}$ | $6(m-1)^3(n-3)^{(2)}$ |
| $jls$ | $(m-1)(n-3)^{(3)}$ | $6(m-2)(n-3)^{(3)}$ | $3(m-2)^{(2)}(n-3)^{(3)}$ | $6(m-3)(n-3)^{(3)}$ | $6(m-3)^{(2)}(n-3)^{(3)}$ | $(m-3)^{(3)}(n-3)^{(3)}$ | $(m-1)^3(n-3)^{(3)}$ |
| Total | $(m-1)(n-1)^3$ | $6(m-2)(n-1)^3$ | $3(m-2)^{(2)}(n-1)^3$ | $6(m-3)(n-1)^3$ | $6(m-3)^{(2)}(n-1)^3$ | $(m-3)^{(3)}(n-1)^3$ | $(m-1)^3(n-1)^3$ |

No new principle is involved in finding these third order moments and so I have not reproduced the algebraic detail. The numbers of terms involved is not without interest and so they are given in Table II. Each cell, however, will need to be split again according to the permutations of the suffices under consideration.

## 10. Product of $k$-functions

In finding the central third moments given in Table I, I found it convenient to derive the crude moments and then reduce them. This meant some algebraic manipulations became rather heavy, with transforming $k$-functions to $K$-functions and conversely. Accordingly, it proved profitable to tabulate products of the Augmented Factorial Monomial Symmetric Functions (AFMSF's). Thus, for example,

$$
\begin{aligned}
(10.1)\qquad [k_3k_1][k_1^2] &= \Big(\sum_{\ell\neq h} k_\ell^{(3)}k_h^{(1)}\Big)\Big(\sum_{u\neq v} k_uk_v\Big)\\
&= 2\sum_{\ell\neq h}(k_\ell^{(3)}k_\ell)(k_h^2) + 2\sum_{\ell\neq h\neq u}(k_\ell^{(3)}k_\ell)(k_h)(k_u)\\
&\quad + 2\sum_{\ell\neq h\neq u}(k_\ell^{(3)})(k_h^2)(k_u) + \sum_{\ell\neq h\neq u\neq v}(k_\ell^{(3)})(k_h)(k_u)(k_v)\\
&= 2[k_4k_2] + \{6[k_3k_2] + 2[k_4k_1]\} + \{6[k_3k_1]\}\\
&\quad + 2[k_4k_1^2] + \{6[k_3k_1^2]\} + 2[k_3k_2k_1] + \{2[k_3k_1^2]\}\\
&\quad + [k_3k_1^3].
\end{aligned}
$$

Table AI of the Appendix gives the expression of the products of all separates of weight 6, 5, 4, 3, 2, in terms of the AFMSF's. This table enables any reduction from crude moments to central moments to be made reasonably quickly.

## 11. Unconditional moments

Expression of the moments in terms of the AFMSF's calls for the minimum of manipulation when we turn from the conditional to the unconditional case. With free multinomial sampling

$$
\begin{aligned}
(11.1)\qquad \mathcal{E}[k_ak_bk_c\cdots] &= \sum_{\ell_1\neq \ell_2\neq \ell_3} k_{\ell_1}^{(a)}k_{\ell_2}^{(b)}k_{\ell_3}^{(c)}\cdots\\
&= K_1^{(a+b+c+\cdots)}\sum_{\ell_1\neq\ell_2\neq\ell_3\neq\cdots} p_{\ell_1}^{*a}p_{\ell_2}^{*b}p_{\ell_3}^{*c}\cdots,
\end{aligned}
$$

with the reduction of the last summation following from the AMSF's of [2]. With Pólya sampling of the type delineated earlier, writing

$$(11.2)\qquad p_\ell^{*[a]_\delta} = p_\ell^*[p_\ell^* + \delta][p_\ell^* + 2\delta]\cdots[p_\ell^* + (a-1)\delta],$$

we have

$$(11.3)\qquad \mathcal{E}[k_ak_bk_c\cdots] = \frac{K_1^{(a+b+c+\cdots)}}{1^{[a+b+c\cdots]_\delta}}\sum_{\ell_1\neq\ell_2\neq\ell_3\cdots} p_{\ell_1}^{*[a]_\delta}p_{\ell_2}^{*[b]_\delta}p_{\ell_3}^{*[c]_\delta}\cdots.$$

Thus, for example, conditionally,

(11.4) $$\mu'_{20} = p^4(m-1)^2(n-1)^2\{[k_4] + [k_2^2]\} + 4p^3(2m-3)(2n-3)[k_3] + \tfrac{1}{2}p^2(3m-4)(3n-4)[k_2].$$

If the contents of the $m \times n$ squares have supposedly arisen from sampling from a multiple Pólya urn scheme, then unconditionally

(11.5)
$$\begin{aligned}\mu'_{20} = {} & p^4(m-1)^2(n-1)^2 \frac{K_1^{(4)}}{1(1+\delta)(1+2\delta)(1+3\delta)} \\ & \left\{\sum_{\ell=1}^{s} p_\ell^*(p_\ell^*+\delta)(p_\ell^*+2\delta)(p_\ell^*+3\delta) + \sum_{\ell \neq h} p_\ell^*(p_\ell^*+\delta)p_h^*(p_h^*+\delta)\right\} \\ & + 4p^3(2m-3)(2n-3)\frac{K_1^{(3)}}{1(1+\delta)(1+2\delta)}\left\{\sum_{\ell=1}^{s} p_\ell^*(p_\ell^*+\delta)(p_\ell^*+2\delta)\right\} \\ & + \frac{1}{2}p^2(3m-4)(3n-4)\frac{K_1^{(2)}}{1(1+\delta)}\left\{\sum_{\ell=1}^{s} p_\ell^*(p_\ell^*+\delta)\right\}.\end{aligned}$$

Here $\delta = 0$ gives the multinomial result, $\delta = -1/N$ gives the "no replacement" result. Generally,

(11.6)
$$\begin{aligned}\mu'_{20} = {} & p^4(m-1)^2(n-1)^2 \frac{K_1^{(4)}}{(1+\delta)(1+2\delta)(1+3\delta)} \\ & \{P_2^2 + \delta(4P_3 + 2P_2) + \delta^2(10P_2 + 1) + 6\delta^3\} \\ & + 4p^3(2m-3)(2n-3)\frac{K_1^{(3)}}{(1+\delta)(1+2\delta)}\{P_3 + 3P_2\delta + 2\delta^2\} \\ & + \frac{1}{2}p^2(3m-4)(3n-4)\frac{K_1^{(2)}}{(1+\delta)}\{P_2 + \delta\}\end{aligned}$$

with

(11.7) $$\mu'_{10} = p^2(m-1)(n-1)\frac{K_1^{(2)}}{1+\delta}\{P_2 + \delta\}.$$

No general simplification appears possible for the central moments, except in the multinomial case. For this latter case, we have

(11.8)
$$\begin{aligned}\mu'_{30} = {} & p^6(m-1)^3(n-1)^3K_1^{(6)}P_2^3 \\ & + 12p^5(m-1)(n-1)(2m-3)(2n-3)K_1^{(5)}P_3P_2 \\ & + \tfrac{3}{2}p^4(m-1)(n-1)(3m-4)(3n-4)K_1^{(4)}P_2^2 \\ & + p^4[6(8m-15)(8n-15) + 8(4m-7)(4n-7)]P_4K_1^{(4)} \\ & + p^3[24(3m-5)(3n-5) + 2(5m-8)(5n-8)]K_1^{(3)}P_3 \\ & + p^2(5m-8)(5n-8)K_1^{(2)}P_2.\end{aligned}$$

For brevity, let

(11.9)
$$\begin{aligned}K_1^{(6)} - 3K_1^{(4)}K_1^{(2)} + 2(K_1^{(2)})^3 &= F, \\ K_1^{(5)} - K_1^{(3)}K_1^{(2)} &= G, \\ K_1^{(4)} - (K_1^{(2)})^2 &= H.\end{aligned}$$

Then

$$
\begin{aligned}
(11.10)\quad \mu_{30} = {} & p^6(m-1)^3(n-1)^3P_2^3F \\
& +p^5(m-1)(n-1)(2m-3)(2n-3)12P_3P_2G \\
& +\tfrac{3}{4}p^4(m-1)(n-1)(3m-4)(3n-4)P_2^2H \\
& +p^4[6(8m-15)(8n-15)+8(4m-7)(4n-7)P_4K_1^{(4)} \\
& +p^3[24(3m-5)(3n-5)+2(5m-8)(5n-8)]P_3K_1^{(3)} \\
& +p^2(5m-8)(5n-8)P_2K_1^{(2)},
\end{aligned}
$$

$$
\begin{aligned}
(11.11)\quad \mu_{21} = {} & p^6(m-1)^3(n-1)^3P_2^2(1-P_2)F \\
& +4p^5(m-1)(n-1)(2m-3)(2n-3)[-3P_2P_3+2P_2^2+P_3]G \\
& +\tfrac{1}{2}p^4(m-1)(n-1)(3m-4)(3n-4)P_2(1-P_2)H \\
& +2p^4(8m-15)(8n-15)K_1^{(4)}(-3P_4+P_2^2+2P_3) \\
& +8p^4(4m-7)(4n-7)K_1^{(4)}[-P_4+P_3] \\
& +8p^3(3m-5)(3n-5)K_1^{(3)}(-P_3+P_2),
\end{aligned}
$$

(11.12)

$$
\begin{aligned}
\mu_{12} = {} & p^6(m-1)^3(n-1)^3P_2(1-P_2)^2F \\
& +4p^5(m-1)(n-1)(2m-3)(2n-3)[3P_2P_3-2P_3-4P_2^2+3P_2]G \\
& +\tfrac{1}{2}p^4(m-1)(n-1)(3m-4)(3n-4)P_2(1-P_2)H \\
& +2p^4(8m-15)(8n-15)K_1^{(4)}(3P_4-2P_2^2-4P_3+2P_2) \\
& +8p^4(4m-7)(4n-7)K_1^{(4)}[P_4-2P_3+P_2] \\
& +p^3[8(3m-5)(3n-5)+2(5m-8)(5n-8)K_1^{(3)}[-P_3+P_2],
\end{aligned}
$$

(11.13)

$$
\begin{aligned}
\mu_{30} = {} & p^6(m-1)^3(n-1)^3(1-P_2)^3F \\
& +12p^5(m-1)(n-1)(2m-3)(2n-3)G(1-P_2)(P_3-2P_2+1) \\
& +\tfrac{3}{2}p^4(m-1)(n-1)(3m-4)(3n-4)(1-P_2)^2H \\
& +6p^4(8m-15)(8n-15)K_1^{(4)}[(1-P_2)^2-(P_4-2P_3+P_2)] \\
& +8p^4(4m-7)(4n-7)K_1^{(4)}[1-3P_2+3P_3-P_4] \\
& +24\,p^3(3m-5)(3n-5)K_1^{(3)}(1-2P_2+P_3) \\
& +2p^3(5m-8)(5n-8)K_1^{(3)}[1-2P_2+P_3] \\
& +p^2(5m-8)(5n-8)(1-P_2)K_1^{(2)}.
\end{aligned}
$$

It remains to note that $F = 40K_1^{(4)} + 64K_1^{(3)} + 8K_1^{(2)}$ with similar reductions for the others.

## 12. Distribution of $S_W$ and $S_B$

In order to reduce $S_W$ or $S_B$ to a diversity index, it was suggested in [1] that for an observed score $S_0$ it would be appropriate to calculate $P\{S \leqq S_0\}$ and use this quantity as a measure of aggregation, or segregation, in diversity. It will however be recognized that this requires the distribution of $S$ and this is difficult to obtain except possibly for the case of one dimension. Such random sampling experiments as we have done indicates that both $S_W$ and $S_B$ have distributions which look like $\chi^2$ (or $\chi$) and we would accordingly suggest that a suitable approximating function to these distributions may be $\chi^2$, keeping the normal approximation for use when $\beta_1$ is very small. Thus, we write

$$\mathcal{E}(S) - A = af, \qquad \operatorname{Var} S = 2a^2f, \qquad \mu_3(S) = 8a^3f, \tag{12.1}$$

so that

$$a = \frac{\mu_3}{4\mu_2}, \qquad f = \frac{8}{\beta_1}, \qquad A = \mathcal{E}(S) - \frac{2\mu_2^2}{\mu_3}. \tag{12.2}$$

This will usually mean that $f$, the degrees of freedom, are fractional, and interpolation into the $\chi^2$ tables will be necessary. On the whole, such sampling results as we have indicate that for reasonable sized chessboards and with $K_1$ not very different from $M = mn$ the $\chi^2$ approximation is adequate. For one dimension as indicated below the start may be fixed at one half below the first possible frequency of $S$ and the first two moments only will then yield values of $a$ and $f$. It will not escape notice that the numerical calculation of $\mu_3$ will be heavy arithmetically and we have in fact written a program for it.

12.1. *One dimension.* The distribution of $S_W$ for five balls of the same color in five boxes arranged in a line is as in Table III.

The momental constants are

$$\begin{aligned} &\mu_1' = 12.8, \qquad \mu_2 = 28.032, \qquad \sigma = 5.294525 \\ &\beta_1 = 1.4288, \qquad \beta_2 = 4.7768. \end{aligned} \tag{12.3}$$

For the $(\beta_1\beta_2)$ point to lie on the Type III line, we should have $2\beta_2 - 3\beta_1 - 6 = 0$, or for given $\beta_1$, we have $\beta_2 = 5.1432$.

The true $(\beta_1\beta_2)$ point accordingly lies in the Type I area but it is not far off the Type III line.

If $A$, $a$, and $f$ have the meanings of the previous section then $A = 3.94136$, $a = 1.5822$, $f = 5.5989$. We use the $\chi^2$ tables and obtain $\chi^2_{0.05} = 11.9832$; thus, the frequency beyond $a\chi^2_{0.05}$ is 183 which is 5.86 per cent. The $\chi^2_{0.01} = 16.1216$; thus the frequency beyond $a\chi^2_{0.01}$ is 33 or 1.056 per cent. This example takes an extreme case and the approximation using $\chi^2_{0.05}$ may be expected to become better as the number of balls and/or boxes increases. For instance, suppose we drop five balls of one color and four balls of another color in five boxes arranged in a line. The distribution of $S_W$ is shown in Table IV.

The momental constants are:

$$\begin{aligned} &\mu_1' = 20.48, \qquad \mu_2 = 43.008, \qquad \sigma = 6.5580, \qquad \beta_1 = 0.7673, \\ &A = 5.50656, \qquad f = 10.42615^+, \qquad a = 1.43614. \end{aligned} \tag{12.4}$$

The $\chi^2_{0.05}$ point cuts off 4.6 per cent from the upper tail and the $\chi^2_{0.01}$, 1.1 per cent.

If $\beta_1$ is of negligible proportions, which implies that $f$, the degrees of freedom of the $\chi^2$ are large, the normal curve may be used. For small degrees of freedom, as above, the use of the normal curve may be misleading. Thus, $1.6449 \times 6.5580 + 20.48 = 31.26$. The actual percentage frequency beyond 31.26 is 7.39 per cent.

TABLE III

DISTRIBUTION OF $S_W$ FOR FIVE BALLS OF THE SAME COLOR IN FIVE BOXES ARRANGED IN A LINE

| $S_W$ | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 | 26 | 28 | 30 | 32 | 34 | 36 | 38 | 40 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 30 | 220 | 520 | 640 | 430 | 540 | 110 | 220 | 212 | 20 | 90 | 20 | 40 | . | 30 | . | . | . | 3 | 3125 |

TABLE IV

DISTRIBUTION OF $S_W$ FOR FIVE BALLS OF ONE COLOR AND FOUR BALLS OF ANOTHER COLOR IN FIVE BOXES ARRANGED IN A LINE

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $S_W$ | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 | 26 | 28 | 30 | 32 | 34 |
| Frequency | 720 | 9420 | 47880 | 127860 | 21560 | 258000 | 270760 | 235480 | 197008 | 155776 | 136650 | 88322 | 65716 | 54500 | 25424 |
| $S_W$ | 36 | 38 | 40 | 42 | 44 | 46 | 48 | 50 | 52 | 54 | 56 | 58 | 60 | 62 | 64 |
| Frequency | 25794 | 16388 | 8710 | 5052 | 4110 | 1884 | 1008 | 924 | 330 | 36 | 126 | 72 | · | · | 9 |

Total 1953125

The complete enumeration of $S$ for two dimensions called for extensive computations, given any realistic set of numbers. Accordingly, we resorted to random sampling. Given $m = 4$, $n = 5$, $k_1 = 10$, $k_2 = 8$, $k_3 = 5$, $k_4 = 4$, (that is, $K_1 = 27$), the 27 points were put down randomly on the $4 \times 5$ chessboard and $S_W$ and $S_B$ calculated. In all, this process was carried out 500 times. The distributions obtained are shown in Table V. The observed value of $\beta_2$ is 3.88. The

TABLE V

DISTRIBUTION OF $S_W$ IN 500 SAMPLES
$m = 4, n = 5, k_1 = 10, k_2 = 8, k_3 = 5, k_4 = 4$

| $S_W$ | Frequency | $S_W$ | Frequency | $S_W$ | Frequency |
|---|---|---|---|---|---|
| 36 | 1 | 88 | 20 | 140 | 1 |
| 38 | — | 90 | 14 | 142 | 2 |
| 40 | — | 92 | 22 | 144 | — |
| 42 | — | 94 | 15 | 146 | — |
| 44 | 2 | 96 | 12 | 148 | — |
| 46 | 4 | 98 | 12 | 150 | 1 |
| 48 | — | 100 | 16 | 152 | 1 |
| 50 | 6 | 102 | 11 | 154 | — |
| 52 | 3 | 104 | 6 | 156 | — |
| 54 | 10 | 106 | 5 | 158 | — |
| 56 | 9 | 108 | 12 | 160 | — |
| 58 | 10 | 110 | 10 | 162 | — |
| 60 | 10 | 112 | 8 | 164 | — |
| 62 | 15 | 114 | 5 | 166 | — |
| 64 | 10 | 116 | 1 | 168 | — |
| 66 | 17 | 118 | 6 | 170 | — |
| 68 | 22 | 120 | 7 | 172 | 1 |
| 70 | 11 | 122 | 6 | 174 | 1 |
| 72 | 21 | 124 | 3 | 176 | 1 |
| 74 | 20 | 126 | 3 | | |
| 76 | 31 | 128 | 1 | Total | 500 |
| 78 | 19 | 130 | 3 | | |
| 80 | 17 | 132 | 1 | | |
| 82 | 12 | 134 | 1 | | |
| 84 | 22 | 136 | 3 | | |
| 86 | 15 | 138 | — | | |

condition for the Type III line is $2\beta_2 - 3\beta_1 - 6 = 0$, which will give for the theoretical $\beta_2$ (if the Type III assumption is justified) a value of 3.87. It is clear that a Type III may be a suitable approximating function, although the $(\beta_1\beta_2)$ point might suggest the inverse of the Type III, that is, the Type V. Using Type III and the $\chi^2$ tables with fractional degrees of freedom, we obtain the values of $S_W$, in both the observed and theoretical cases, which should cut off 50 per cent, 5 per cent, and 1 per cent from the right tail of the distribution (Table VI). These values are then used on the actual sampling distribution of $S_W$ to obtain the true percentage tail content. The results are given in Table VII. For the case we are considering $20(= 5 \times 4)$ cells and 27 objects the agreement between the estimated percentage points and the true 5 per cent points would appear

TABLE VI

MOMENTAL CONSTANTS FOR DISTRIBUTION OF $S_W$

| | $\mu_1$ | $\mu_2$ | $\mu_3$ | $\beta_1$ |
|---|---|---|---|---|
| Observed | 85.424 | 486.4042 | 7550.01 | 0.495 |
| Theoretical | 85.44 | 457.4336 | 7457.88 | 0.581 |

TABLE VII

PERCENTAGE POINTS FOR $S_W$ USING CONSTANTS OF TABLE VI AND THE TYPE III DISTRIBUTION

| | | | | Nominal points | | |
|---|---|---|---|---|---|---|
| | $A$ | $a$ | $f$ | 50% | 5% | 1% |
| Observed | 22.75 | 3.881 | 16.151 | 49.8 | 4.2 | 1.2 |
| Theoretical | 29.33 | 4.076 | 13.767 | 49.8 | 4.2 | 1.2 |

reasonably good. With a larger chessboard, and thus a lessening of the part played by the edge effects, the agreement should be closer.

For $S_B$ we have the following distribution in the 500 samples shown in Tables VIII, IX, and X.

The values of the momental constants are in reasonably close agreement. Using the assumption that $S_B/2$ is distributed as $a\chi^2$ the nominal 50 per cent, 5 per cent, and 2.5 per cent points were found in each case and the actual percentages were calculated for these points from the sampling distribution.

12.2. *Two dimensions (Pielou's examples)*. Pielou ([4], p. 180) gives two examples of spatial patterns which she has created in order to illustrate specific points. We use her data here to illustrate the application of the various criteria proposed. Two species only are considered. We denote these by $A$ and $C$. The areas were gridded by us into $10 \times 8$ cells and the count per cell was determined. The counts for Pattern 1 are shown in Table XI.

We note that $A$ and $C$ together have a random pattern; $A$ and $C$ are segregated (that is, tend not to occur together). We have as observed moments

$$\begin{aligned} &S_T = 2246, \quad &&\bar{S}_T = 2325.015, \quad &&\operatorname{Var} S_T = 28835.906275,\\ (12.5)\quad &S_W = 1664, &&\bar{S}_W = 1349.775, &&\operatorname{Var} S_W = 12551.642875,\\ &S_B = 582, &&\bar{S}_B = 975.24, &&\operatorname{Var} S_B = 7906.4874,\\ & && &&\operatorname{Cov} S_W S_B = 4188.888. \end{aligned}$$

Calculation of the third moments show the skewness to be negligible. Consequently from normal tables,

(12.6)

$$P\{S_T \leqq S_{T_0}\} = 0.309, \qquad P\{S_W \leqq S_{W_0}\} = 0.998, \qquad P\{S_B \leqq S_{B_0}\} = 0.231.$$

TABLE VIII

Distribution of $S_B/2$ in 500 Samples
$m = 4, n = 5, k_1 = 10, k_2 = 8, k_3 = 5, k_4 = 4$

| $S_B/2$ | Frequency | $S_B/2$ | Frequency | $S_B/2$ | Frequency | $S_B/2$ | Frequency |
|---|---|---|---|---|---|---|---|
| 70 | 1 | 113 | 5 | 156 | 3 | 199 | — |
| 71 | — | 114 | 7 | 157 | 2 | 200 | — |
| 72 | — | 115 | 1 | 158 | 1 | 201 | — |
| 73 | — | 116 | 12 | 159 | 8 | 202 | — |
| 74 | — | 117 | 9 | 160 | — | 203 | — |
| 75 | — | 118 | 5 | 161 | 2 | 204 | — |
| 76 | 1 | 119 | 10 | 162 | 4 | 205 | — |
| 77 | — | 120 | 7 | 163 | 1 | 206 | — |
| 78 | 1 | 121 | 10 | 164 | 4 | 207 | — |
| 79 | 2 | 122 | 7 | 165 | 4 | 208 | 1 |
| 80 | 2 | 123 | 8 | 166 | 1 | 209 | — |
| 81 | 2 | 124 | 7 | 167 | 7 | 210 | — |
| 82 | 2 | 125 | 6 | 168 | 2 | 211 | — |
| 83 | 4 | 126 | 11 | 169 | 3 | 212 | — |
| 84 | 1 | 127 | 6 | 170 | — | 213 | — |
| 85 | — | 128 | 10 | 171 | 2 | 214 | — |
| 86 | 3 | 129 | 8 | 172 | 1 | 215 | — |
| 87 | 2 | 130 | 6 | 173 | 1 | 216 | 1 |
| 88 | 3 | 131 | 4 | 174 | 1 | 217 | — |
| 89 | 2 | 132 | 5 | 175 | — | 218 | — |
| 90 | 2 | 133 | 5 | 176 | 2 | 219 | — |
| 91 | 3 | 134 | 5 | 177 | — | 220 | — |
| 92 | 4 | 135 | 5 | 178 | — | 221 | — |
| 93 | 7 | 136 | 3 | 179 | — | 222 | — |
| 94 | 9 | 137 | 8 | 180 | 2 | 223 | — |
| 95 | 2 | 138 | 8 | 181 | 1 | 224 | — |
| 96 | 4 | 139 | 1 | 182 | — | 225 | — |
| 97 | 6 | 140 | 5 | 183 | 2 | 226 | — |
| 98 | 6 | 141 | 4 | 184 | 1 | 227 | — |
| 99 | 2 | 142 | 6 | 185 | — | 228 | — |
| 100 | 8 | 143 | 4 | 186 | 1 | 229 | — |
| 101 | 4 | 144 | 8 | 187 | 1 | 230 | — |
| 102 | 6 | 145 | 3 | 188 | 2 | 231 | — |
| 103 | 4 | 146 | 4 | 189 | 1 | 232 | — |
| 104 | 6 | 147 | 8 | 190 | 1 | 233 | — |
| 105 | 10 | 148 | 3 | 191 | — | 234 | — |
| 106 | 12 | 149 | 3 | 192 | — | 235 | 1 |
| 107 | 9 | 150 | 5 | 193 | — | 236 | — |
| 108 | 6 | 151 | 11 | 194 | — | 237 | 1 |
| 109 | 10 | 152 | 3 | 195 | — | | |
| 110 | 5 | 153 | 6 | 196 | — | Total | 500 |
| 111 | 12 | 154 | 7 | 197 | — | | |
| 112 | 11 | 155 | 4 | 198 | — | | |

TABLE IX

Distribution of $S_B/2$ in 500 Samples
Momental Constants

| | $\mu_1'$ | $\mu_2$ | $\mu_3$ | $\beta_1$ |
|---|---|---|---|---|
| Observed | 126.146 | 673.5287 | 11516.5453 | 0.43 |
| Theoretical | 125.76 | 655.7792 | 11485.8965 | 0.47 |

TABLE X

Percentage Points for $S_B/2$ Using the Constants of Table IX and the Type III Distribution

| | Mean | S.D. | 50% | Nominal 5% | 2.5% |
|---|---|---|---|---|---|
| Observed | 126.146 | 25.95 | 49.0% | 3.8% | 2.0% |
| Theoretical | 125.760 | 25.61 | 50.6% | 4.0% | 2.2% |

TABLE XI

Counts for Pattern 1

| A's | | | | | | | | | | C's | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 2 | 0 | 0 | 0 | 1 | 2 | 1 | 1 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 0 | 4 | 2 | 1 | 2 | 2 | 0 | 0 | 2 | 1 | 3 | 0 | 0 | 0 | 1 | 3 |
| 1 | 2 | 2 | 0 | 0 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 2 |
| 2 | 0 | 2 | 0 | 2 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 1 |
| 1 | 0 | 0 | 1 | 2 | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 4 | 1 |
| 2 | 2 | 0 | 2 | 2 | 3 | 1 | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 1 | 2 | 1 | 0 | 1 | 2 | 2 | 0 | 1 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total A's: 86 | | | | | | | | | | Total C's: 36 | | | | | | | | | |

Since the correlation between the within and between scores is high ($\rho = +0.4205$) and $S_W$ is large, it would seem appropriate to modify $S_B$ by calculating the quantity

$$\frac{1}{(1-\rho^2)^{1/2}}\left(\frac{S_B - \bar{S}_B}{\sigma_{S_B}} - \rho\,\frac{S_W - \bar{S}_W}{\sigma_{S_W}}\right) = S_B^* \qquad \text{(say)}. \tag{12.7}$$

From unit normal tables $S_B^*$ cuts off 0.017 from the left tail. It is clear that if the $86 + 36 = 122$ are put down randomly on the $10 \times 8$ chessboard, then the ideal randomness score will give a value of 0.5 for $P\{S_T \leqq S_{T_0}\}$. If the observed value $S_{T_0}$ is large, then this will imply clumping of the combined values; if it is small, then we would suspect there was regularity in the positioning of the $A$'s and the

$C$'s. The fact that $S_W$ is large indicates that the $A$'s and the $C$'s tend to be clumped among themselves. A glance at the counts shows this indeed to be the case. The small value of $S_B^*$ indicates that $A$ tends not to occur with $C$, which is again the case.

TABLE XII

COUNTS FOR PATTERN 2

| A's | | | | | | | | | | C's | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 1 | 2 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| 0 | 0 | 0 | 3 | 3 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 0 |
| 0 | 0 | 0 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 |
| 0 | 2 | 2 | 0 | 0 | 0 | 2 | 1 | 2 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 0 | 3 | 2 | 0 | 0 | 0 | 0 | 3 | 0 | 3 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| 0 | 1 | 0 | 0 | 0 | 0 | 2 | 2 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| Total A's: 58 | | | | | | | | | | Total C's: 30 | | | | | | | | | |

The counts for Pattern 2 are shown in Table XII. Here we note a two species population in which the plants have clumped patterns but the species are unsegregated (that is, tend to occur together).

We have as observed moments

$$
\begin{aligned}
&S_T = 2216, & \bar{S}_T &= 1205.82, & \operatorname{Var} S_T &= 12020.2071,\\
(12.8)\qquad &S_W = 1132, & \bar{S}_W &= 657.72, & \operatorname{Var} S_W &= 4869.2856,\\
&S_B = 1084, & \bar{S}_B &= 548.10, & \operatorname{Var} S_B &= 3776.5395,\\
& & & & \operatorname{Cov} S_B S_W &= 1687.1910.
\end{aligned}
$$

The indices for $S_T$, $S_W$, $S_B$ and $S_{B^*}$ are all unity to five decimal places. Thus, neither the species, considered together or separately have a random pattern and $A$ and $C$ are not random with respect to their juxtaposition.

Table XIII shows the counts for Pattern 3, which is a random arrangement of

TABLE XIII

COUNTS FOR PATTERN 3

| A's | | | | | | | | | | C's | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 0 | 1 | 2 | 1 | 2 | 1 | 0 | 1 | 1 | 1 | 0 | 2 | 0 | 2 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 2 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 1 |
| 1 | 0 | 3 | 4 | 0 | 2 | 1 | 0 | 2 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 |
| 0 | 2 | 1 | 1 | 2 | 0 | 0 | 1 | 1 | 3 | 0 | 2 | 0 | 0 | 2 | 1 | 0 | 1 | 1 | 0 |
| 0 | 4 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 2 | 0 | 1 | 2 | 0 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 0 |
| 1 | 2 | 0 | 3 | 1 | 2 | 0 | 3 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |

Pattern 1. In fact the 86 $A$'s and 36 $C$'s of Pattern 1 were put down randomly using a random number table.

We have the observed

$$S_T = 2134,\quad S_W = 1170,\quad S_B = 964, \tag{12.9}$$

with indices $S_T$: 0.130, $S_W$: 0.054, and $S_B$: 0.450, and

$$S_W(A) = 996,\quad S_W(C) = 174, \tag{12.10}$$

observed separately, with indices $S_W(A)$: 0.072 and $S_W(C)$: 0.244.

Although the letters were put down randomly there is a suspicion of the antithesis of clumping as evidenced by the low value of the index for $S_W$. We break up $S_W$ into the two parts of the contribution for $A$ and the contribution for $C$, calculate the separate means and variances and calculate an index for each. The indications are that there is a suspicion of nonrandomness among those "random" numbers used to put down the $A$'s.

## 13. Discussion

The spatial pattern of points representing objects belonging to different categories has been examined in a series of papers relating to the distribution of the chromosomes in the human cell in mitosis. The material algebra for these investigations was given in Barton and David [5]. The essence of the method is to recognize that the plotted centromeres (that is, the central points of the chromosomes) will most likely not be spatially random, but the numbers attached to the chromosomes might be. Accordingly, it was suggested that the variance of two like numbers be compared with the overall variance.

There is no doubt that this randomization test could be extended and applied to this present problem, but there is an essential difference. Given several species of plants ($A, B, C, D, \cdots$), it is hypothesized that the spatial points representing these plants are randomly distributed over the area. (In the chromosome case we were reasonably certain that they were not, and we were principally interested in whether the like pairs tended to lie too close together.) Under the hypothesis alternate to randomness we ask do the plants of the same species tend to cluster together (possibly in many clusters) and is there any tendency for there to be a segregation of the different species?

It is recognized that the method used—that of gridding the area—may lead to a loss of information, and certainly the optimum number of cells to use for the chessboard is a matter for investigation. Intuitively, one feels that an average of one observation per cell should be aimed at, but only the specification of suitable alternate hypotheses could decide this. Further, and more importantly, the scoring for the four cells surrounding each node is flexible. For example, if there are $t_{ij\ell}$ plants of the $\ell$th kind and $t_{ijh}$ of the $h$th kind in the four cells surrounding the $(ij)$th node, then the contribution towards aggregation has been taken as $t_{ij\ell}^{(2)} + t_{ijh}^{(2)}$ and the contribution towards segregation as $t_{ij\ell} \times t_{ijh}$. A different system, suggested in [1] might be to exclude contributions which are only joined by a diagonal line. For example, consider the sets of four cells in Figure 1. Under the system of scoring of this present paper, we would count 6, 3, 1, 1, respectively. But if we adopt a nondiagonal system of scoring, we would count 4, 2, 1, 0. Again, only the specification of an alternate hypothesis would make a decision possible as to which was the optimum procedure.

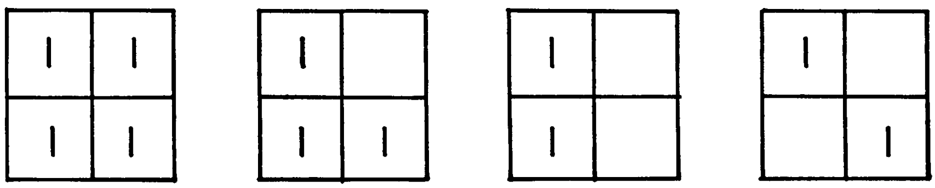

FIGURE 1
Sets of four cells

## 14. Previous work

As mentioned in the last paragraph, the basic work on one aspect of spatial randomization tests was given in [1]. A suggestion of gridding and counting joins will be found in Pielou [4], although not the scoring system. The first of these present papers was done independently of Dr. Pielou's work which leads one to the hope that we may be on the right lines. A paper by Mandel and [6] concerned with the clustering of cases of a disease, also using "joins," appeared in a recent journal, but seems different from the present approach.

◇ ◇ ◇ ◇ ◇

The material discussed here arose from problems presented to the writer by the scientific staff of the Pacific-Southwest Range and Experiment Station, U.S.D.A.

◇ ◇ ◇ ◇ ◇

## APPENDIX

Table AI gives the products of all separates of weight for the $k$-functions in terms of the AFMSF's.

Table AII gives the $K$-functions in terms of the AFMSF's and *vice versa*. Equations (A.1) and (A.2) illustrate the notation:

$$[k_3 k_2 k_1] = \sum_{a \neq b \neq c} k_a^{(3)} k_b^{(2)} k_c^{(1)}, \tag{A.1}$$

$$K_r = \sum_a k_a^{(r)}. \tag{A.2}$$

TABLE AI

PRODUCTS OF SEPARATES UP TO WEIGHT 6 FOR $k$-FUNCTIONS IN TERMS OF THE AUGMENTED MONOMIAL SYMMETRIC FUNCTIONS

| | $[k_6]$ | $[k_5k_1]$ | $[k_4k_2]$ | $[k_4k_1^2]$ | $[k_3^2]$ | $[k_3k_2k_1]$ | $[k_3k_1^3]$ | $[k_2^3]$ | $[k_2^2k_1^2]$ | $[k_2k_1^4]$ | $[k_1^6]$ | $[k_5]$ | $[k_4k_1]$ | $[k_3k_2]$ | $[k_3k_1^2]$ | $[k_2^2k_1]$ | $[k_2k_1^3]$ | $[k_1^5]$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $[k_6]$ | 1 | | | | | | | | | | | | | | | | | |
| $[k_5][k_1]$ | 1 | 1 | | | | | | | | | | 5 | | | | | | |
| $[k_4k_1][k_1]$ | | 1 | 1 | 1 | | | | | | | | | 5 | | | | | |
| $[k_3k_2][k_1]$ | | | 1 | | 1 | 1 | | | | | | | | 5 | | | | |
| $[k_3k_1^2][k_1]$ | | | | 1 | | 2 | 1 | | | | | | | | 5 | | | |
| $[k_2^2k_1][k_1]$ | | | | | | 2 | | 1 | 1 | | | | | | | 5 | | |
| $[k_2k_1^3][k_1]$ | | | | | | | 1 | | 3 | 1 | | | | | | | 5 | |
| $[k_1^5][k_1]$ | | | | | | | | | | 5 | 1 | | | | | | | 5 |
| $[k_4k_2]$ | 1 | | 1 | | | | | | | | | 8 | | | | | | |
| $[k_3k_1][k_2]$ | | 1 | | | 1 | 1 | | | | | | | 6 | 2 | | | | |
| $[k_2^2][k_2]$ | | | 2 | | | | | 1 | | | | | | 8 | | | | |
| $[k_2k_1^2][k_2]$ | | | | 1 | | 2 | | | 1 | | | | | | 4 | 4 | | |
| $[k_1^4][k_2]$ | | | | | | | 4 | | | 1 | | | | | | | 8 | |
| $[k_4][k_1^2]$ | | 2 | | 1 | | | | | | | | | 8 | | | | | |
| $[k_3k_1][k_1^2]$ | | | 2 | 2 | | 2 | 1 | | | | | | 2 | 6 | 8 | | | |
| $[k_2^2][k_1^2]$ | | | | | 2 | 4 | | | 1 | | | | | 8 | | 8 | | |
| $[k_2k_1^2][k_1^2]$ | | | | | | 4 | 2 | 2 | 4 | 1 | | | | | 4 | 12 | 8 | |
| $[k_1^4][k_1^2]$ | | | | | | | | | 12 | 8 | 1 | | | | | | 24 | 8 |
| $[k_3]^2$ | 1 | | | | 1 | | | | | | | 9 | | | | | | |
| $[k_2k_1][k_3]$ | | 1 | 1 | | | 1 | | | | | | | 6 | 3 | | | | |
| $[k_1^3][k_3]$ | | | | 3 | | | 1 | | | | | | | | 9 | | | |
| $[k_2k_1]^2$ | | | 1 | 1 | 1 | 2 | | 1 | 1 | | | | 1 | 8 | 4 | 5 | | |
| $[k_2k_1][k_1^3]$ | | | | | | 6 | 3 | | 3 | 1 | | | | | 6 | 12 | 9 | |
| $[k_1^3]^2$ | | | | | | | | 6 | 18 | 9 | 1 | | | | | 18 | 36 | 9 |
| $[k_4][k_1]^2$ | 1 | 2 | 1 | 1 | | | | | | | | 9 | 9 | | | | | |
| $[k_3][k_2][k_1]$ | 1 | 1 | 1 | | 1 | 1 | | | | | | 11 | 6 | 5 | | | | |
| $[k_3k_1][k_1][k_1]$ | | 1 | 2 | 2 | 1 | 3 | 1 | | | | | | 9 | 9 | 9 | | | |
| $[k_3][k_1^2][k_1]$ | | 2 | 2 | 3 | | 2 | 1 | | | | | | 16 | 6 | 11 | | | |
| $[k_2]^3$ | 1 | | 3 | | | | | 1 | | | | 12 | | 12 | | | | |
| $[k_2]^2[k_1^2]$ | | 2 | | 1 | 2 | 4 | | | 1 | | | | 16 | 8 | 4 | 8 | | |
| $[k_2k_1][k_2][k_1]$ | | 1 | 2 | 1 | 1 | 3 | | 1 | 1 | | | | 9 | 13 | 4 | 7 | | |
| $[k_2][k_1^3][k_1]$ | | | | 3 | | 6 | 4 | | 3 | 1 | | | | | 21 | 12 | 11 | |
| $[k_2][k_1^2]$ | | | 4 | 4 | | 8 | 4 | 2 | 4 | 1 | | | 4 | 20 | 28 | 20 | 12 | |

TABLE AI (Continued)

| | $[k_4]$ | $[k_3k_1]$ | $[k_2^2]$ | $[k_2k_1^2]$ | $[k_1^4]$ | $[k_3]$ | $[k_2k_1]$ | $[k_1^3]$ | $[k_2]$ | $[k_1^2]$ | $[k_1]$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $[k_6]$ | | | | | | | | | | | |
| $[k_5][k_1]$ | | | | | | | | | | | |
| $[k_4k_1][k_1]$ | | | | | | | | | | | |
| $[k_3k_2][k_1]$ | | | | | | | | | | | |
| $[k_3k_1^2][k_1]$ | | | | | | | | | | | |
| $[k_2^2k_1][k_1]$ | | | | | | | | | | | |
| $[k_2k_1^3][k_1]$ | | | | | | | | | | | |
| $[k_1^5][k_1]$ | | | | | | | | | | | |
| $[k_4][k_2]$ | 12 | | | | | | | | | | |
| $[k_3k_1][k_2]$ | | 6 | | | | | | | | | |
| $[k_2^2][k_2]$ | | | 4 | | | | | | | | |
| $[k_2k_1^2][k_2]$ | | | | 2 | | | | | | | |
| $[k_1^4][k_2]$ | | | | | | | | | | | |
| $[k_4][k_1^2]$ | | | | | | | | | | | |
| $[k_3k_1][k_1^2]$ | | 6 | | | | | | | | | |
| $[k_2^2][k_1^2]$ | | | 8 | | | | | | | | |
| $[k_2k_1^2][k_1^2]$ | | | | 10 | | | | | | | |
| $[k_1^4][k_1^2]$ | | | | | 12 | | | | | | |
| $[k_3]^2$ | 18 | | | | | 6 | | | | | |
| $[k_2k_1][k_3]$ | | 6 | | | | | | | | | |
| $[k_1^3][k_3]$ | | | | | | | | | | | |
| $[k_2k_1]^2$ | | 4 | 6 | 2 | | | 2 | | | | |
| $[k_2k_1][k_1^3]$ | | | | 12 | | | | | | | |
| $[k_1^3]^2$ | | | | 18 | 18 | | | 6 | | | |
| $[k_4][k_1]^2$ | 16 | | | | | | | | | | |
| $[k_3][k_2][k_1]$ | 30 | 6 | | | | 18 | | | | | |
| $[k_3k_1][k_1][k_1]$ | | 16 | | | | | | | | | |
| $[k_3][k_1^2][k_1]$ | | 24 | | | | | | | | | |
| $[k_2]^3$ | 38 | | 6 | | | 32 | | | 4 | | |
| $[k_2]^2[k_1^2]$ | | 28 | 8 | 2 | | | 8 | | | | |
| $[k_2k_1][k_2][k_1]$ | | 18 | 10 | 2 | | | 6 | | | | |
| $[k_2][k_1^3][k_1]$ | | | | 24 | | | | | | | |
| $[k_2][k_1^2]^2$ | | 20 | 16 | 34 | | | 16 | | | | |

TABLE AI (Continued)

| | $[k_6]$ | $[k_5k_1]$ | $[k_4k_2]$ | $[k_4k_1^2]$ | $[k_3^2]$ | $[k_3k_2k_1]$ | $[k_3k_1^3]$ | $[k_2^3]$ | $[k_2^2k_1^2]$ | $[k_2k_1^4]$ | $[k_1^6]$ | $[k_5]$ | $[k_4k_1]$ | $[k_3k_2]$ | $[k_3k_1^2]$ | $[k_2^2k_1]$ | $[k_2k_1^3]$ | $[k_1^5]$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $[k_2k_1][k_1^2][k_1]$ | | | 2 | 2 | 2 | 10 | 3 | 2 | 5 | 1 | | | 2 | 20 | 18 | 26 | 11 | |
| $[k_2k_1^2][k_1]^2$ | | | | 1 | | 6 | 2 | 2 | 5 | 1 | | | | | 6 | 18 | 9 | |
| $[k_1^4][k_1]^2$ | | | | | | | 4 | | 12 | 9 | 1 | | | | | | 36 | 9 |
| $[k_1^3][k_1^2][k_1]$ | | | | | | 12 | 6 | 6 | 24 | 11 | 1 | | | | 12 | 54 | 66 | 11 |
| $[k_1^2]^3$ | | | | | 4 | 24 | 8 | 8 | 30 | 12 | 1 | | | 24 | 24 | 96 | 84 | 12 |
| $[k_3][k_1]^3$ | 1 | 3 | 3 | 3 | 1 | 3 | 1 | | | | | 12 | 24 | 12 | 12 | | | |
| $[k_2]^2[k_1]^2$ | 1 | 2 | 3 | 1 | 2 | 4 | | 1 | 1 | | | 13 | 17 | 23 | 4 | 9 | | |
| $[k_2k_1][k_1]^3$ | | 1 | 4 | 3 | 3 | 13 | 3 | 3 | 6 | 1 | | 7 | 19 | 22 | 24 | 29 | 12 | |
| $[k_2][k_1^2][k_1]^2$ | | 2 | 4 | 5 | 2 | 12 | 4 | 2 | 5 | 1 | | | 22 | 29 | 36 | 30 | 13 | |
| $[k_1^2]^2[k_1]^2$ | | | 4 | 4 | 4 | 32 | 12 | 10 | 34 | 13 | 1 | | 4 | 49 | 56 | 126 | 104 | 13 |
| $[k_1^3][k_1]^3$ | | | | 3 | | 18 | 10 | 6 | 27 | 12 | 1 | | | | 36 | 72 | 84 | 12 |
| $[k_2][k_1]^4$ | 1 | 4 | 7 | 6 | 4 | 16 | 4 | 3 | 6 | 1 | | 14 | 42 | 56 | 42 | 42 | 14 | |
| $[k_1^2][k_1]^4$ | | 2 | 8 | 9 | 6 | 44 | 16 | 12 | 39 | 14 | 1 | | 28 | 84 | 98 | 168 | 126 | 14 |
| $[k_1]^6$ | 1 | 6 | 15 | 15 | 10 | 60 | 20 | 15 | 45 | 15 | 1 | 15 | 75 | 150 | 150 | 225 | 150 | 15 |
| $[k_5]$ | | | | | | | | | | | | 1 | | | | | | |
| $[k_4][k_1]$ | | | | | | | | | | | | 1 | 1 | | | | | |
| $[k_3][k_2]$ | | | | | | | | | | | | 1 | | 1 | | | | |
| $[k_3][k_1^2]$ | | | | | | | | | | | | | 2 | | 1 | | | |
| $[k_3k_1][k_1]$ | | | | | | | | | | | | | 1 | 1 | 1 | | | |
| $[k_3][k_1]^2$ | | | | | | | | | | | | 1 | 2 | 1 | 1 | | | |
| $[k_2^2][k_1]$ | | | | | | | | | | | | | | 2 | | 1 | | |
| $[k_2k_1][k_2]$ | | | | | | | | | | | | | 1 | 1 | | 1 | | |
| $[k_2]^2[k_1]$ | | | | | | | | | | | | 1 | 1 | 2 | | 1 | | |
| $[k_2][k_1^3]$ | | | | | | | | | | | | | | | 3 | | 1 | |
| $[k_2k_1][k_1^2]$ | | | | | | | | | | | | | | 2 | 2 | 2 | 1 | |
| $[k_2k_1^2][k_1]$ | | | | | | | | | | | | | | | 1 | 2 | 1 | |
| $[k_2k_1][k_1]^2$ | | | | | | | | | | | | | 1 | 3 | 2 | 3 | 1 | |
| $[k_3][k_1]^3$ | | | | | | | | | | | | 1 | 3 | 4 | 3 | 3 | 1 | |
| $[k_1^4][k_1]$ | | | | | | | | | | | | | | | | | 4 | 1 |
| $[k_1^3][k_1^2]$ | | | | | | | | | | | | | | | | 6 | 6 | 1 |
| $[k_1^3][k_1]^2$ | | | | | | | | | | | | | | | 3 | 6 | 7 | 1 |
| $[k_1^2]^2[k_1]$ | | | | | | | | | | | | | | 4 | 7 | 4 | 3 | |
| $[k_1^2][k_1]^3$ | | | | | | | | | | | | | 2 | 6 | 7 | 12 | 9 | 1 |
| $[k_1]^5$ | | | | | | | | | | | | 1 | 5 | 10 | 10 | 15 | 10 | 1 |

TABLE AI (Continued)

| | $[k_4]$ | $[k_3k_1]$ | $[k_2^2]$ | $[k_2k_1^2]$ | $[k_1^4]$ | $[k_3]$ | $[k_2k_1]$ | $[k_1^3]$ | $[k_2]$ | $[k_1^2]$ | $[k_1]$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $[k_2k_1][k_1^2][k_1]$ | | 12 | 20 | 28 | | | 12 | | | | |
| $[k_2k_1^2][k_1]^2$ | | | | 16 | | | | | | | |
| $[k_1^4][k_1]^2$ | | | | | 16 | | | | | | |
| $[k_1^3][k_1^2][k_1]$ | | | | 66 | 30 | | | 18 | | | |
| $[k_1^2]^3$ | | 8 | 36 | 120 | 30 | | 24 | 32 | | 4 | |
| $[k_3][k_1]^3$ | 37 | 37 | | | | 27 | | | | | |
| $[k_2]^2[k_1]^2$ | 46 | 33 | 20 | 2 | | 46 | 12 | | 8 | | |
| $[k_2k_1][k_1]^3$ | 28 | 37 | 9 | 37 | | | 27 | | | | |
| $[k_2][k_1^2][k_1]^2$ | | 59 | 26 | 44 | | | 34 | | | | |
| $[k_1^2]^2[k_1]^2$ | | 33 | 66 | 291 | 46 | | 58 | 46 | | 8 | |
| $[k_1^3][k_1]^3$ | | | | 111 | 37 | | | 27 | | | |
| $[k_2][k_1]^4$ | 55 | 110 | 55 | 55 | | 65 | 65 | | 16 | | |
| $[k_1^2][k_1]^4$ | | 110 | 110 | 380 | 55 | | 130 | 65 | [illegible] | 16 | |
| $[k_1]^6$ | 65 | 260 | 195 | 495 | 65 | 90 | 270 | 90 | 31 | 31 | 1 |
| $[k_5]$ | | | | | | | | | | | |
| $[k_4][k_1]$ | 4 | | | | | | | | | | |
| $[k_3][k_2]$ | 6 | | | | | 6 | | | | | |
| $[k_3][k_1^2]$ | | 6 | | | | | | | | | |
| $[k_3k_1][k_1]$ | | 4 | | | | | | | | | |
| $[k_3][k_1]^2$ | 7 | 7 | | | | 9 | | | | | |
| $[k_2^2][k_1]$ | | | 4 | | | | | | | | |
| $[k_2k_1][k_2]$ | | 4 | 2 | | | | 2 | | | | |
| $[k_2]^2[k_1]$ | 8 | 4 | 4 | | | 14 | 2 | | | | |
| $[k_2][k_1^3]$ | | | | 6 | | | | | | | |
| $[k_2k_1][k_1^2]$ | | 2 | 4 | 6 | | | 4 | | | | |
| $[k_2k_1^2][k_1]$ | | | | 4 | | | | | | | |
| $[k_2k_1][k_1]^2$ | | 7 | 7 | 7 | | | 9 | | | | |
| $[k_2][k_1]^3$ | 9 | 18 | 9 | 9 | | 19 | 19 | | 8 | | |
| $[k_1^4][k_1]$ | | | | | 4 | | | | | | |
| $[k_1^3][k_1^2]$ | | | | 12 | 6 | | | 1 | | | |
| $[k_1^3][k_1]^2$ | | | | 21 | 7 | | | 9 | | | |
| $[k_1^2]^2[k_1]$ | | 4 | 12 | 26 | 2 | | 16 | 8 | | 4 | |
| $[k_1^2][k_1]^3$ | | 18 | 18 | 45 | 9 | | 38 | 19 | | 8 | |
| $[k_1]^5$ | 10 | 40 | 30 | 60 | | 25 | 75 | 25 | 15 | 15 | 1 |

TABLE AI (Continued)

| | $[k_4]$ | $[k_3k_1]$ | $[k_2^2]$ | $[k_2k_1^2]$ | $[k_1^4]$ | $[k_3]$ | $[k_2k_1]$ | $[k_1^3]$ | $[k_2]$ | $[k_1^2]$ | $[k_1]$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $[k_4]$ | 1 | | | | | | | | | | |
| $[k_3][k_1]$ | 1 | 1 | | | | 3 | | | | | |
| $[k_2]^2$ | 1 | | 1 | | | 4 | | | 2 | | |
| $[k_3][k_1^2]$ | | 2 | | 1 | | | 4 | | | | |
| $[k_2k_1][k_1]$ | | 1 | 1 | 1 | | | 3 | | | | |
| $[k_2][k_1]^2$ | 1 | 2 | 1 | 1 | | 5 | 5 | | 4 | | |
| $[k_1^3][k_1]$ | | | | 3 | 1 | | | 3 | | | |
| $[k_1^2]^2$ | | | 2 | 4 | 1 | | 4 | 4 | | 2 | |
| $[k_1^2][k_1]^2$ | | 2 | 2 | 5 | 1 | | 8 | 5 | | 2 | |
| $[k_1]^4$ | 1 | 4 | 3 | 6 | 1 | 6 | 18 | 6 | 7 | 7 | 1 |
| $[k_3]$ | | | | | | 1 | | | | | |
| $[k_2][k_1]$ | | | | | | 1 | 1 | | 2 | | |
| $[k_1^2][k_1]$ | | | | | | | 2 | 1 | | 2 | |
| $[k_1]^3$ | | | | | | 1 | 3 | 1 | 3 | 3 | 1 |
| $[k_2]$ | | | | | | | | | 1 | | |
| $[k_1]^2$ | | | | | | | | | 1 | 1 | 1 |

## TABLE AII

$K$-FUNCTIONS IN TERMS OF THE AFMSF'S AND *Vice Versa*

| Weight 6 | $[k_6]$ | $[k_5k_1]$ | $[k_4k_2]$ | $[k_4k_1^2]$ | $[k_3^2]$ | $[k_3k_2k_1]$ | $[k_3k_1^3]$ | $[k_2^3]$ | $[k_2^2k_1^2]$ | $[k_2k_1^4]$ | $[k_1^6]$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $K_6$ | **1** | −1 | −1 | 2 | −1 | 2 | −6 | 2 | −6 | 24 | −120 |
| $K_5(K_1 - 5)$ | 1 | **1** | | −2 | | −1 | 6 | | 4 | −24 | 144 |
| $K_4(K_2 - 12) - 8K_5$ | 1 | | **1** | −1 | | −1 | 3 | −3 | 5 | −18 | 90 |
| $K_4(K_1 - 4)^{(2)}$ | 1 | 2 | 1 | **1** | | | −3 | | −1 | 12 | −90 |
| $K_3(K_3 - 6) - 9(K_5 + 2K_4)$ | 1 | | | | **1** | −1 | 2 | | 2 | −8 | 40 |
| $(K_3(K_2 - 6) - 6K_4)(K_1 - 5)$ | 1 | 1 | 1 | | 1 | **1** | −3 | | −4 | 20 | −120 |
| $K_3(K_1 - 3)^{(3)}$ | 1 | 3 | 3 | 3 | 1 | 3 | **1** | | | −4 | 40 |
| $K_2(K_2 - 2)(K_2 - 4) - 12K_3(K_2 - 2) + 40(K_4 + K_3)$ | 1 | | 3 | | | | | **1** | −1 | 3 | −15 |
| $(K_2(K_2 - 2) - 4K_3)(K_1 - 4)^{(2)}$ | 1 | 2 | 3 | 1 | 2 | 4 | | 1 | **1** | −6 | 45 |
| $K_2(K_1 - 4)^{(4)}$ | 1 | 4 | 7 | 16 | 4 | 16 | 4 | 3 | 6 | **1** | −15 |
| $K_1^{(6)}$ | 1 | 6 | 15 | 15 | 10 | 60 | 20 | 15 | 45 | 15 | **1** |

To express $K$-functions in terms of the AFMSF's, read horizontally up to and including the heavy type diagonal; for example,

$$(K_3(K_2 - 6) - 6K_4)(K_1 - 5) = [k_6] + [k_5k_1] + [k_4k_2] + [k_3^2] + [k_3k_2k_1]. \tag{A.3}$$

To express the AFMSF's in terms of the $K$-functions, read vertically downward including the heavy type diagonal; for example,

(A.4)

$$\begin{aligned}[k_3k_2k_1] = {} & 2K_6 - K_5(K_1 - 5) - [K_4(K_2 - 12) - 8K_5] \\ & - [K_3(K_3 - 6) - 9(K_5 + 2K_4)] + [K_3(K_2 - 6) - 6K_4][K_1 - 5].\end{aligned}$$

## REFERENCES

[1] F. N. DAVID, "Measurement of diversity," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, Vol. 1, 1972, pp. 631–648.

[2] F. N. DAVID and M. G. KENDALL, "Tables of symmetric functions," *Biometrika*, Vol. 36 (1950), pp. 431–449.

[3] F. N. DAVID, M. G. KENDALL, and D. E. BARTON, *Symmetric Function and Allied Tables*, Cambridge, Cambridge University Press, 1967.

[4] E. C. PIELOU, *An Introduction to Mathematical Ecology*, New York, Wiley-Interscience, 1969.

[5] D. E. BARTON and F. N. DAVID, "Randomisation bases for multivariate tests," *Proc. I.S.I., 32nd Session* (1961), pp. 44–51.

[6] N. MANTEL and J. C. BAILAR, "A class of permutational and multinomial tests," *Biometrics*, Vol. 26 (1971), pp. 687–700.