This, in general, will not be the case, for on most soils there is a block difference. In this particular test the ground used had been previously filled in with well mixed soil.    The efficiency for the analysis given in [1] relative to the randomized complete blocks was less than 1.00.

This paper and the previous one show what a long tedious procedure is necessary to analyze the data, when the design does not follow the rules for the construction of the lattice, triple lattice, etc.    The complexity of these methods stresses the importance, to those designing experiments, of not deviating from the established design if the most information is to be secured from the data with simple calculations.

### REFERENCES

[1]  BOYD HARSHBARGER, "On the analysis of a certain six-by-six four-group lattice design," *Annals of Math. Stat.*, Vol. 15, No. 3 (1944), pp. 307–320.
[2]  F. YATES, "A new method of arranging variety trials involving a large number of varieties," *Journal Agri. Science*, Vol. 26 (1936), pp. 424–455.
[3]  F. YATES, "The recovery of inter-block information in three dimensional lattice," *Ann. Eugenics*, Vol. 9 (1939), pp. 136–156.
[4]  F. YATES, "The recovery of inter-block information in balanced incomplete block designs," *Ann. Eugenics*, Vol. 10 (1940), pp. 317–325.
[5]  F. YATES, "Lattice squares," *Journal Agri. Science*, Vol. 3 (1940), pp. 672–687.
[6]  G. M. COX, R. C. ECKHARDT, AND W. G. COCHRAN, "The analysis of lattice and triple lattice experiments in corn varietal tests," *Iowa Agri. Exp. Sta. Res. Bul.*, Vol. 281 (1940).

# FURTHER REMARKS ON LINKAGE THEORY IN MENDELIAN HEREDITY

BY HILDA GEIRINGER

*Wheaton College*

In the following an explicit formula for the distribution of genotypes in case of three Mendelian characters will be given [formula (5)].    The complete discussion of the case $m = 3$ suggests a supplement (as stated in the last paragraph of this paper) to the general limit theorem dealing with $m$ characters.

In an earlier paper[1] recurrence formulae have been derived which furnish the distribution of genotypes in the $n$th generation if the distribution in the $(n - 1)$th generation and the "linkage distribution" (l.d.) are known.    It was also shown how to "integrate" this system of difference equations so as to determine the distribution in the $n$th generation directly from that in the 0th generation. This last method, though straightforward, requires however in each particular case quite a few operations.

In case $m$, the number of Mendelian characters, equals two, an explicit formula for the problem in question had been known.    Denote by $p(x_1, x_2)$,

---

[1] HILDA GEIRINGER,  *Annals of Math. Stat.*  Vol. 15 (1944), pp.25–57.   The notation in the present Note will be the same as in this paper.

$(x_1, x_2 = 1, 2, \cdots k)$, the "distribution of transmitted genes" in the original, 0th, generation, by $p^{(n)}(x_1, x_2)$ that in the $n$th generation and by $c$ the "crossover probability" (c.p.).    Then the simple formula holds:[2]

$$(1) \qquad p^{(n)}(x_1, x_2) = (1 - c)^n p(x_1, x_2) + [1 - (1 - c)^n] p_1(x_1) p_2(x_2).$$

This may also be written:

$$(1') \qquad p^{(n)}(x_1, x_2) = p_1(x_1) p_2(x_2) + (1 - c)^n [p(x_1, x_2) - p_1(x_1) p_2(x_2)],$$

where $p_i(x_i)$ are the marginal distributions derived from $p(x_1, x_2)$.  (1') shows that, if in case of independence of the original distribution, $p(x_1, x_2) = p_1(x_1) p_2(x_2)$ then $p^{(n)}(x_1, x_2) = p(x_1, x_2)$ for every $n$.   The same is true for arbitrary $p(x_1, x_2)$ if $c = 0$.   Otherwise, if $c > 0$ the second term to the right in (1') tends towards zero as $n \to \infty$ and the well known limit theorem results.

In case $m = 3$, a remarkably elegant explicit formula exists[3] which may be deduced from the author's general theory.   In this case the l.d. is completely equivalent to the three c.p.'s $c_{12}, c_{23}, c_{13}$.   The $c_{ij}$ are probabilities with sum $\leqq$ 2, and for which the triangular relation

$$(2) \qquad\qquad\qquad c_{ij} + c_{jk} \geqq c_{ik}$$

holds.   If $l(\epsilon_1, \epsilon_2, \epsilon_3)$ $(\epsilon_i = 0, 1)$ denotes the eight values of the l.d. we have (see quot. [1], p. 32) $l(000) = l(111), l(100) = l(011), l(010) = l(101), l(001) = l(110)$, hence three independent values only.   We may introduce

$$(3) \quad \begin{aligned} 2l(000) &= v(000) = v_0, \ 2l(100) = v(100) \equiv v_1, \ 2l(010) = v(010) \equiv v_2 \\ 2l(001) &= v(001) \equiv v_3; \quad v_0 + v_1 + v_2 + v_3 = 1. \end{aligned}$$

It follows easily that

$$(4) \qquad\qquad c_{ij} = v_i + v_j, \qquad (i \neq j, \ i, j = 1, 2, 3).$$

The original distribution $p(x_1, x_2, x_3)$ has marginal distributions $p_{ij}(x_i, x_j)$, $p_i(x_i)$.   These values will be denoted briefly by $p_{123}, p_{12}, p_{23}, p_{13}, p_1, p_2, p_3$ respectively.   Writing in an analogous way $p^{(n)}(x_1 x_2 x_3) = p_{123}^{(n)}$ the new formula is the following:

$$(5) \quad \begin{aligned} p_{123}^{(n)} &= p_1 p_2 p_3 + [(v_0 + v_1)^n - v_0^n](p_1 p_{23} - p_1 p_2 p_3) + [(v_0 + v_2)^n - v_0^n](p_2 p_{13} \\ &\quad - p_1 p_2 p_3) + [(v_0 + v_3)^n - v_0^n](p_3 p_{12} - p_1 p_2 p_3) + v_0^n(p_{123} - p_1 p_2 p_3). \end{aligned}$$

This useful formula permits to compute readily $p_{123}^{(n)}$ for every $n$.   In terms of the $c_{ij}$, writing

$$(6) \qquad\qquad d_{ij} = 1 - c_{ij}, \qquad v_0 = 1 - \tfrac{1}{2}(c_{12} + c_{23} + c_{13}),$$

it reads

$$(5') \quad p_{123}^{(n)} = p_1 p_2 p_3 + (d_{23}^n - v_0^n)(p_1 p_{23} - p_1 p_2 p_3) + \cdot + \cdot \ v_0^n(p_{123} - p_1 p_2 p_3).$$

---

[2] H. S. Jennings, *Genetics*, Vol. 12 (1917) pp. 97–154.

[3] Professor Felix Bernstein called this author's attention to the biologically interesting case $m = 3$.

In these formulae the role of independence of the original distribution is clearly seen: If $p_{ij} = p_i p_j$ and $p_{123} = p_1 p_2 p_3$ then $p_{123}^{(n)} = p_{123}$ for every $n$ and every l.d. The same holds for every $n$ and every $p_{123}$ if $v_0 = 1$, which implies that all $c_{ij}$ be zero. If in (5') all $d_{ij} < 1$, hence all $c_{ij} > 0$ the limit theorem $\lim_{n \to \infty} p_{123}^{(n)} = p_1 p_2 p_3$ results. $c_{ij} > 0$ means that complete linkage between any two genes is excluded. If, on the other hand, e.g. $v_0 > 0$, $v_1 > 0$, $v_0 + v_1 \equiv d_{23} = 1$, $c_{23} = 0$, hence $v_0 < 1$, $v_2 = v_3 = 0$ we get $p_{123}^{(n)} \to p_1 p_{23}$. If $c_{23} = c_{12} = 0$ the triangular relation (2) shows that $c_{13} = 0$ too, a case considered above.

It should be noticed that (5) is, of course, in agreement with the author's equation (41) in quot. [1]. It only has to be observed,—an obvious fact not mentioned in my earlier paper,—that in the former setup the sum of all the $\alpha^{(n)}$ for every fixed $m$ equals one. Thus for $m = 3$:

(7) $$\alpha_{123}^{(n)} + \alpha_{1,23}^{(n)} + \alpha_{2,13}^{(n)} + \alpha_{3,12}^{(n)} + \alpha_{1,2,3}^{(n)} = 1, \text{ (for every } n),$$

and

(8) $$\alpha_{123}^{(n)} = v_0^n, \qquad \alpha_{1,23}^{(n)} = (v_0 + v_1)^n - v_0^n = d_{23}^n - v_0^n.$$
$$\alpha_{2,13}^{(n)} = (v_0 + v_2)^n - v_0^n = d_{13}^n - v_0^n.$$
$$\alpha_{3,12}^{(n)} = (v_0 + v_3)^n - v_0^n = d_{12}^n - v_0^n.$$

The preceeding complete discussion of the case $m = 3$ suggests a remark concerning the general case of $m$ characters. In my earlier paper the influence on the main limit theorem of certain ways of degeneration of the l.d. had not been explicitly considered. In the following we shall use the $v$-distribution which is a little shorter to write than the l.d. $l(\epsilon_1, \epsilon_2, \cdots \epsilon_m)$. The $v$-distribution contains only $2^{m-1}$ values with sum one, defined in a way similar to (3). The main limit theorem ([1], theorem II, p. 42) states in our present notation that

(9) $$\lim_{n \to \infty} p_{12\cdots m}^{(n)} = p_1 p_2 \cdots p_m,$$

if "complete linkage" between any group of genes is excluded. That implies that not only $v_0 \equiv v(0, 0, \cdots 0) = 1$ must be excluded but even $v_{ij\ldots k}(0, \cdots 0) = 1$, where this last probability denotes a marginal distribution of the $v$-distribution of an order $\geqq 2$. To assure this it is necessary and sufficient that no $v_{ij}(0, 0) = 1$, or no $d_{ij} \equiv v_{ij}(0, 0) = 1$, or no $c_{ij} = 0$. Hence (9) holds if and only if no $c_{ij} = 0$. If this condition is not satisfied the l.d. degenerates in various ways and the limit theorem is to be modified accordingly. If, in particular, $v_0 = 1$, all $c_{ij} = 0$, and $p_{12\cdots m}^{(n)} = p_{12\ldots m}$ for every $n$.

Between these two extreme cases ("no $c_{ij} = 0$", "all $c_{ij} = 0$") are the different possibilities of $r < m$ groups of completely linked-characters (see [1] p. 36, iv)). Consider e.g. $m = 7$ and $v_{1234}(0000) = 1$, $v_{567}(000) = 1$ (this is realized if $v(0000000) > 0$, $v(0000111) > 0$ with sum of these two numbers equal to one) then $\lim_{n \to \infty} p_{12\cdots 7}^{(n)} = p_{1234} p_{567}$. Here the four characters 1, 2, 3, 4 act as one character and $p_{1234}^{(n)} = p_{1234}$ for every $n$. Also $p_{567}^{(n)} = p_{567}$. Or if, for $m = 6$, $d_{12} = d_{34} = d_{56} = 1$ (realized if $v(000000) > 0$, $v(110000) > 0$, $v(001100) > 0$, $v(000011) > 0$, with

the sum of these four values equal to one) then $p_{12\ldots6}^{(n)} \to p_{12}p_{34}p_{56}$ . If however for $m = 6$ merely $d_{12} = d_{34} = 1$ (realized if, in a notation analogous to (3), $v_0$ , $v_5$ , $v_6$ , $v_{56}$ , $v_{12}$ , $v_{34}$ , $v_{125}$ , $v_{126}$ are the only non-zero values of the l.d.) then $p_{12\ldots6}^{(n)} \to p_{12}p_{34}p_5p_6$ .

In general, with a proof which consists in a modification of the reasoning (p. 41), of my earlier paper, we may state the following complement to the main limit theorem (9): *If the l.d. is such that $r < m$ disjoint groups $G_1$ , $G_2$ ,$\cdots$ $G_r$ of completely linked characters exist, i.e. such that within each group no crossover takes place, each group containing as many of the m numbers as compatible with the definition but not less than two, and all groups together containing $s \leqq m$ of the m elements, then, as $n \to \infty$, $p_{12\ldots m}^{(n)}$ converges towards the product of those marginal distributions (of the original generation) which correspond to these groups multiplied by the marginal distributions of order one of the remaining free elements which are not contained in any such group. In a formula:*

$$(10) \qquad \lim_{n \to \infty} p_{G_1, G_2, \cdots G_r, \gamma_{s+1}, \gamma_{s+2}, \cdots \gamma_m} = p_{G_1} p_{G_2} \cdots p_{G_r} p_{\gamma_{s+1}} p_{\gamma_{s+2}} \cdots p_{\gamma_m} .$$

We may also characterize these linked groups of maximum size by stating that while within each group no crossover takes place there must be at least one c.p. $\neq$ 0 among any two such groups and at least one among any group and any free element. It may however be noted that if there is one c.p. $> 0$ among two groups of complete linkage (or among a group and a free element) then all c.p.'s among these two groups are different from zero. In fact, it follows by repeated use of the triangular relation (2) that if one c.p. among two disjoint groups of complete linkage is zero, all of them are zero. If, e.g., $(1, 2, 3)$ and $(5, 6, 8)$ are two groups of complete linkage, i.e. $v_{123}(000) = 1$ and $v_{568}(000) = 1$ and if besides $c_{15} = 0$, then $v_{123568}(000000) = 1$ and these six elements form a group of complete linkage.

It may be noticed that the above statement of the generalized limit theorem becomes simpler and more elegant by counting "free elements" as groups. It might then run as follows: *If $G_1$ , $G_2$ , $\cdots$ $G_t(t \leqq m)$ are the maximal groups of completely linked characters, then, under the hypotheses of the earlier paper, the gene distribution in successive generations approaches a limit in which the original (marginal) probabilities within each group $G_i$ are preserved and genes and sets of genes from different groups are independently distributed.*

---

# ON THE DEFINITION OF DISTANCE IN THE THEORY OF THE GENE

By Hilda Geiringer

*Wheaton College*

In several letters to this author Dr. I. M. H. Etherington of the University of Edinburgh has raised questions concerning the author's definition of "distance" proposed in Section 10 of her paper on Mendelian heredity,[1] comparing it with

---

[1] *Annals of Math. Stat.*, Vol. 15 (1944), pp. 25–57.