ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
Hugo Steinhaus Jubilee Volume
X (1969)

REGINA C. ELANDT-JOHNSON (Chapel Hill)

APPLICATION OF OCCUPANCY AND ORDERING THEORY IN GENETICS OF AUTOPOLYPLOIDS

1. Introduction and summary.

The purpose of this paper is to present and summarize some results concerning the number of possible genotypes and the probabilities of their occurrence through generations in polysomic inheritance. The general pattern for genotype with several loci and multiple alleles at each locus in $2m$-autopolyploid breeding populations is considered. Using combinatorial analysis and generating functions, formulae for the number of different genotypes, the number of different gametes produced by the same genotype, the number of genotypes producing the same kind of gametes, etc. are obtained.

These results are useful in evaluating the expected frequencies of occurrence of gametes (segregation distributions) and the proportions of genotypes in random mating populations.

Although some of these results are known or have been discussed by other authors (Fisher [3], Geiringer [5]-[10], Moran [12]), it is still, perhaps, worthwhile to revise them, find more general and simpler models which can be applied in the population genetics of polyploids under fairly general conditions.

An extensive discussion of gamete distributions under random chromosome and random chromatide segregation in panmictic populations has been presented in a series of papers by Hilda Geiringer [5]-[10]. She emphasized that maternal and paternal influence on the frequency of gamete production can be of significant importance. She introduced in her models some kind of "chromosome linkage" which results in tendency of maternal (and paternal) genes to segregate more or less closely together. The genotype distribution in the offspring is affected by this assumption. Moran ([12], p. 39) expressed some doubts whether the distinction between the maternal and paternal origin is of biological significance.

Without loss of generality this assumption is omitted in this paper; this also has the advantage of making the presentation rather simpler and more comprehensive. If under a certain hypothesis on the mode of inheritance on finds the segregation probabilities, then using the general model presented in section 4.1, the genotype distribution in the offspring can be obtained.

To avoid some ambiguity in defining terms like: genotype, chromosome and chromatid segregation, etc. we first discuss briefly these terms and their definitions.

2. Some basic ideas and definitions.

Let us consider a bisexual, $2m$-autopolyploid population. Suppose that we are interested in r loci, denoted by $a_1, a_2, \ldots, a_r$, and that s_h randomly segregating alleles are assigned to the locus a_h. We denote these alleles by $a_{h1}, a_{h2}, \ldots, a_{hs_h}$ $(h = 1, 2, \ldots, r)$. The zygote of a $2m$-ploid individual can be represented as a set of elements $a_{ht}^{w_{ht}}$, where $h = 1, 2, \ldots, r$ denotes the locus number, $t = 1, 2, \ldots, s_h$ denotes the t-th allele at the a_h-th locus; $w_{ht} = 0, 1, 2, \ldots, 2m$ denotes the number of alleles of the type a_{ht}, and $\sum_{t=1}^{s_h} w_{ht} = 2m$ at each given locus a_h for any set of w's. More explicitly, the genic pattern for the zygote can be formally written

$$(2.1) \qquad \left(a_{11}^{w_{11}} \ldots a_{1s_1}^{w_{1s_1}} \| a_{21}^{w_{21}} \ldots a_{2s_2}^{w_{2s_2}} \| \ldots \| a_{r1}^{w_{r1}} \ldots a_{rs_r}^{w_{rs_r}}\right) = \prod_{h=1}^{r} \prod_{t=1}^{s_h} a_{ht}^{w_{ht}},$$

with the condition $\sum_{t=1}^{s_h} w_{ht} = 2m$, $h = 1, 2, \ldots, r$.

Similarly, the genic pattern for the gamete can be represented as

$$(2.2) \qquad \left(a_{11}^{u_{11}} \ldots a_{1s_1}^{u_{1s_1}} \| \ldots \| a_{r1}^{u_{r1}} \ldots a_{rs_r}^{u_{rs_r}}\right) = \prod_{h=1}^{r} \prod_{t=1}^{s_h} a_{ht}^{u_{ht}},$$

with $\sum_{t=1}^{s_h} u_{ht} = m$.

Let us first consider the definition of *genotype*. If by genotype we understand simply an individual's genic constitution, the terms: "zygote" and "genotype" are interchangeable. If the genotype does not only describe the genic formula of an individual, but also the way it breeds, the "zygote" and "genotype" do not necessarily have the same meaning. For example, in the case of two linked loci in diploids, the zygotes $a_{11}a_{21}/a_{12}a_{22}$ and $a_{11}a_{22}/a_{12}a_{21}$ have the same genic pattern, but they are different when regarded as "segregating genotypes". To distinguish between these two meanings we introduce the terms: genotype in "broader

sense" as equivalent to zygote pattern (2.1), and genotype in "narrower sense", when the breeding process (production of gametes) is taken into account.

With each "segregating genotype" we can combine the probabilities of occurrence of a series of gametes (segregation distribution). The problem of evaluating the segregation probabilities in polyploids is much more complicated than in diploids and needs special attention. Here we restrict ourselves to two limiting cases of segregation distributions, that is:

a) *random chromosome segregation* which corresponds to the model of forming the gamete by choosing a set of m chromosomes at random;

b) *random chromatid segregation*, that is, when each of the $2m$ chromosomes divides into two chromatids, and then a random selection of m out of $4m$ chromatids is made.

The numbers of different kinds of genotypes and gametes depend only on whether one or another process of segregation occurs and does not depend on whether the process is random or not. But the segregation probabilities depend on the frequency of occurrence of different *modes* of segregation and on whether the segregation process is random or not. Only random segregation processes are considered here.

In evaluating the numbers of different genotypes and gametes we apply combinatorial analysis, using generating functions. An important problem in combinatorial analysis is that of finding in how many different ways n objects can be assigned to k boxes. The objects can be distinguishable or not, the boxes can be ordered (numbered) or not. The excellent book by MacMahon [11] can be used as a reference to this subject.

Let us first discuss the simplest situation of *one* locus.

I. ONE LOCUS WITH s ALLELES IN $2m$-PLOID CELLS

3. Genotype constitution of $2m$-ploid population.

3.1. The numbers of different genotypes and gametes. We consider one locus, a, with s alleles: $a_1, a_2, \ldots, a_s$. The chromosome pattern for a zygote in $2m$-ploid germ cell is

$$(a_1^{w_1} a_2^{w_2} \ldots a_s^{w_s}), \tag{3.1}$$

where $w_t = 0, 1, 2, \ldots, 2m$; $t = 1, 2, \ldots, s$ and $\sum_{t=1}^{s} w_t = 2m$ for any set of w's.

The question arises: how many different zygotes of pattern (3.1), N_z, can be obtained under *chromosome segregation*? The model is as follows: The number of s different *genes* corresponds to s *ordered boxes*. Each gene can appear at most once in each of the $2m$ chromosomes.

Since the number of chromosomes is immaterial, $2m$ *chromosomes* can be treated as $2m$ *indistinguishable objects*. The distribution of genes in the chromosomes corresponds to assigning to each of the numbers $1, 2, \ldots, s$ (ordered boxes), $2m$ *or fewer* chromosomes (indistinguishable objects). The number of different ways in which this can happen is the coefficient of x^{2m} in the expansion of

$$A_1(x) = (1+x+x^2+\ldots+x^{2m})^s \tag{3.2}$$

and is equal to

$$N_z = \binom{s+2m-1}{2m} = \binom{s+2m-1}{s-1}. \tag{3.3}$$

Replacing $2m$ by m we obtain the number of different gametes

$$N_g = \binom{s+m-1}{m}. \tag{3.4}$$

These results are known (see, for example, Geiringer [8], [10]).

If we use for zygote the *chromatid pattern*, that is

$$(a_1^{2w_1} a_2^{2w_2} \ldots a_s^{2w_s}), \tag{3.5}$$

where $\sum_{t=1}^{s} 2w_t = 4m$, then for chromatid segregation the generating function (3.2) remains the same, and so the numbers of genotypes and gametes are again given by (3.3) and (3.4) respectively.

3.2. The number of genotypes having exactly $q \leqslant s$ different genes. The genotypes can be of *homogenic* type; a_t^{2m}, $t = 1, 2, \ldots, s$; of digenic type: $a_{t_1}^{w_1} a_{t_2}^{w_2}$, where $w_1, w_2 = 1, 2, \ldots, (2m-1)$, $w_1+w_2 = 2m$, $t_1, t_2 = 1, 2, \ldots, s$ and $t_1 \neq t_2$; and generally, of *q-genic* type: $a_{t_1}^{w_1} a_{t_2}^{w_2} \ldots a_{t_q}^{w_q}$, where $w_\nu = 1, 2, \ldots [2m-(q-1)], \nu = 1, 2, \ldots, q$, $\sum_{\nu=1}^{q} w_\nu = 2m$, $t_\nu = 1, 2, \ldots, s$ and $t_{\nu'} \neq t_\nu$, for each $\nu, \nu' = 1, 2, \ldots, q$.

The number of genotypes having *exactly* q different genes is equivalent to the number of ways of selecting first sets of q different genes; this can be done in $\binom{s}{q}$ ways. For each set, q chromosomes must be assigned, and the remaining $(2m-q)$ will be distributed randomly in

$$\binom{q+(2m-q)-1}{q-1} = \binom{2m-1}{q-1} \tag{3.6}$$

ways (using (3.3)). Therefore, the *exact* number of *q-genic* types is

$$N_{q\text{-genic}} = \binom{s}{q}\binom{2m-1}{q-1}. \tag{3.7}$$

EXAMPLE 1. Let $s = 5$; $2m = 4$; $N_z = \binom{8}{4} = 70$ different genotypes.

Types	Number of genotypes
homogenic	$\binom{5}{1}\binom{3}{0} = 5$
digenic	$\binom{5}{2}\binom{3}{1} = 30$
trigenic	$\binom{5}{3}\binom{3}{2} = 30$
tetragenic	$\binom{5}{4}\binom{3}{3} = \frac{5}{70}$

3.3. The number of different gametes which can be produced by a given genotype $(a_1^{w_1} \dots a_s^{w_s})$. The given genotype $(a_1^{n_1} \dots a_s^{n_s})$, where $\sum_{t=1}^{s} w_t = 2m$, can produce different gametes of the type $(a_1^{u_1} \dots a_s^{u_s})$, where $\sum_{t=1}^{s} u_t = m$. It is obvious that $0 \leqslant u_t \leqslant w_t$, $t = 1, 2, \dots, s$.

We want to know the number, $N_{g|z}$, of different gametes, which can be produced by the same zygote $(a^{w_1} \dots a_s^{w_s})$.

(i) under *chromosome segregation*, the mathematical model is as follows: there is a set of $2m$ elements, where w_1 elements are of type a_1 (indistinguishable), w_2 of type a_2, and generally, w_t of type $a_t (t = 1, 2, \dots, s)$. The number of different subsets of type $(a_1^{u_1} \dots a_s^{u_s})$ which can be obtained from the set $(a_1^{w_1} \dots a_s^{w_s})$ is equal to the coefficient of x^m in the product

$$(3.8) \qquad A_2(x) = \prod_{t=1}^{s} (1 + x + x^2 + \dots + x^{w_t}) = (1-x)^{-s} \prod_{t=1}^{s} (1 - x^{w_t+1}).$$

Applying the principle of inclusion and exclusion, the coefficient of x^m in (3.8) is

$$(3.9) \qquad N_{g|z} = \binom{s+m-1}{s-1} - \sum_{t=1}^{s} \binom{s+m-w_t-2}{s-1} + \sum_{t \neq t'} \sum \binom{s+m-w_t-w_{t'}-3}{s-1} -$$
$$- \sum_{t \neq t' \neq t''} \sum \sum \binom{s+m-w_t-w_{t'}-w_{t''}-4}{s-1} + \dots \text{ etc.},$$

where the summations are defined by the conditions that $(s+m-w_t-2)$, $(s+m-w_t-w_{t'}-3)$, etc.,... are positive.

We notice that if in some population the gametic output consists of N_g different gametes, then the number of genotypes is N_g^2. But only N_z (defined in (3.3)) have different genic patterns; some patterns are repeated. The quantity $N_{g|z}$ calculated from (3.9) gives also the number

of different gametes which could be *components* of a given zygote. The gametes can be paired according to the rule

$$(a_1^{u_1} \dots a_s^{u_s}) \times (a_s^{u_1'} \dots a_1^{u_s'}) = (a_1^{w_1} \dots a_s^{w_s}),$$

where $\sum_{t=1}^{s} u_t = \sum_{t=1}^{s} u_t' = m$ and $u_t + u_t' = w_t$, $t = 1, 2, \dots, s$.

If we denote the $N_{g|z}$ for the l-th zygote ($l = 1, 2, \dots, N_z$) by $N_{g|z(l)}$, then $N_{g|z(l)}$ given by the formula (3.9) gives also the number of genotypes having the same patterns (repeated genotypes) among all N_g^2 genotypes. Thus, we have

$$\sum_{l=1}^{N_z} N_{g|z(l)} = N_g^2. \tag{3.10}$$

Also

$$P(l) = N_g^{-2} N_{g|z(l)}, \quad l = 1, 2, \dots, N_z \tag{3.11}$$

defines the distribution of different genotypes in the population (of the total size N_g^2).

EXAMPLE 2. Let $s = 3$, $2m = 4$ (tetraploids with three alleles). There are $N_g = \binom{4}{2} = 6$ different possible gametes, i.e. a_1^2, a_2^2, a_3^2, a_1a_2, a_1a_3, a_2a_3, and $N_z = \binom{6}{4} = 15$ different kind of genotypes. With this gametic output, the distribution of $6^2 = 36$ genotypes is

Type of genotype	Genotype constitution						
Homogenic:	a_1^4	a_2^4	a_3^4				
$N_{g\|z(l)}$	1	1	1				
Digenic	$a_1^3a_2$	$a_1^3a_3$	$a_2^3a_1$	$a_2^3a_3$	$a_3^3a_1$	$a_3^3a_2$	
$N_{g\|z(l)}$	2	2	2	2	2	2	2
Digenic duplex	$a_1^2a_2^2$	$a_1^2a_3^2$	$a_2^2a_3^2$				
$N_{g\|z(l)}$	3	3	3				
Trigenic:	$a_1^2a_2a_3$	$a_1a_2^2a_3$	$a_1a_2a_3^2$				
$N_{g\|z(l)}$	4	4	4				

(ii) Under *chromatid segregation* formula (3.9) can be applied with $w_t, w_{t'}, w_{t''}$, etc. replaced by $2w_t, 2w_{t'}, 2w_{t''}$, etc. respectively.

EXAMPLE 3. Let $s = 3$, $2m = 8$, $w_1 = 4$, $w_2 = 3$, $w_3 = 1$. (i.e. the zygote is $(a_1^4 a_2^3 a_3)$). The numbers of different gametes which can be produced by this zygote under *chromosome* segregation is $N_{g|z} = \binom{6}{2} - \left[\binom{1}{2} + \binom{2}{2} + \binom{4}{2}\right] = 8$. And these gametes are:

$$(a_1 a_2^2 a_3), (a_1 a_2^3), (a_1^2 a_2 a_3), (a_1^2 a_2^2), (a_1^3 a_2), (a_1^3 a_3), (a_2^3 a_3), (a_1^4).$$

Under chromatid segregation $N^*_{g|z} = \binom{6}{2} - \binom{3}{2} = 12$. The additional 4 gametes are: $(a_1 a_2 a_3^2), (a_1^2 a_3^2), (a_2^2 a_3^2), (a_2^4)$.

3.4. The number of different genotypes which could produce a given gamete $(a_1^{u_1} \dots a_s^{u_s})$. If we "fixe" the subset $(a_1^{u_1} \dots a_s^{u_s})$, where $\sum_{t=1}^{s} u_t = m$, then the problem is reduced to that of section 3.3, where u_t should be now replaced by $w_t - u_t$, and w_t by $2m - w_t$, when the *chromosome segregation* takes place.

Under *chromatid segregation* the situation is a little bit more complicated. We now replace w_t by $2w_t$ and $2m$ by $4m$. Since the subset $(a_1^{u_1} \dots a_s^{u_s})$, with $\sum_{t=1}^{s} u_t = m$, has been fixed, we have now to select subsets $(a_1^{2w_1 - u_1} \dots$ $\dots a_s^{2w_s - u_s})$, with $\sum_{t=1}^{s} (2w_t - u_t) = 3m$, where each $(4m - u_t)$ objects in each subset $(t = 1, 2, \dots, s)$ are indistinguishable, but subject to the additional conditions that

$$u_t + (2w_t - u_t) \quad \text{is even for} \quad t = 1, 2, \dots, s.$$

The number of different ways in which this can be done, is the coefficient of $x^{(3m - \text{no. of odd } u_t\text{'s})}$ in

$$A_3(x) = (1-x^2)^{-s} \prod_{u_t'=1}^{s} (1 - x^{4m - u_t' + 1}), \tag{3.12}$$

where $u_t' = u_t$ if u_t is even, and $u_t' = u_t + 1$ if u_t is odd.

EXAMPLE 4. Let $s = 3$, $2m = 8$, $u_1 = 2$, $u_2 = 1$, $u_3 = 1$. (i.e. the gamete is of the form $(a_1^2 a_2 a_3)$). The number of different genotypes which can produce the gamete $(a_1^2 a_2 a_3)$ under *chromosome segregation*, $N_{z|g}$, is $\binom{6}{2} + [0 + \dots + 0] = 15$. And they are:

$$(a_1^2 a_2 a_3^5), (a_1^2 a_2^2 a_3^4), (a_1^2 a_2^3 a_3^3), (a_1^2 a_2^4 a_3^2), (a_1^2 a_2^5 a_3),$$
$$(a_1^3 a_2 a_3^4), (a_1^3 a_2^2 a_3^3), (a_1^3 a_2^3 a_3^2), (a_1^3 a_2^4 a_3),$$
$$(a_1^4 a_2 a_3^3), (a_1^4 a_2^2 a_3^2), (a_1^4 a_2^3 a_3),$$
$$(a_1^5 a_2 a_3^2), (a_1^5 a_2^2 a_3),$$
$$(a_1^6 a_2 a_3).$$

Under *chromatid segregation* we have $3m$-no. of odd u_t's $= 10$. We have to find the coefficient of x^{10} in $(1-x^2)^{-3}(1-x^{15})^3$ which reduces to finding the coefficient of z^5 in $(1-z)^{-3}$, and this is $\binom{3+5-1}{5} = 21$. The additional 6 genotypes under chromatid segregation are: $(a_1 a_2 a_3^6)$, $(a_1 a_2^2 a_3^5)$, $(a_1 a_2^3 a_3^4)$, $(a_1 a_2^4 a_3^3)$, $(a_1 a_2^5 a_3^2)$, $(a_1 a_2^6 a_3)$.

4. Segregation distributions.

4.1. Some definitions and notations. Let us denote the gamete $(a_1^{u_1} \dots a_s^{u_s})$ briefly by γ_i. The complete set of all possible gametes at one locus with s alleles in $2m$-ploids can be simply written as the $1 \times N_g$ vector

$$\boldsymbol{\gamma}' = (\gamma_1, \gamma_2, \dots, \gamma_{N_g}) \tag{4.1}$$

and can be also called the *gametic output vector*. Each genotype can be represented as

$$\Gamma_{ij} = \gamma_i \gamma_j, \quad i, j = 1, 2, \dots, N_g, \tag{4.2}$$

where γ_i is the maternal and γ_j the paternal contribution to the zygote Γ_{ij}.

Random mating here corresponds to the mathematical relationship

$$\Pr(\Gamma_{ij} = \gamma_i \gamma_j) = \Pr(\gamma_i) \Pr(\gamma_j). \tag{4.3}$$

The expected result of random mating, where the gametic output is given by (4.1) is a population $\boldsymbol{\Gamma}$ of genotypes Γ_{ij} which can be represented in the form of the $N_g \times N_g$ matrix

$$\boldsymbol{\Gamma} = \{\Gamma_{ij}\}. \tag{4.4}$$

If we do not distinguish between the maternal and paternal effects on the genotype, we have $\Gamma_{ij} = \Gamma_{ji}$; also some other genotypes $\Gamma_{i'j'} (i \neq i', j \neq j')$ might have the same genic pattern as Γ_{ij} as has been pointed out in section 3.3.

Let us denote the conditional probability of occurrence of the gamete γ_k among those produced by the given genotype Γ_{ij} (segregation probability) by $c_{k(ij)}$, that is

$$\Pr(\gamma_k | \Gamma_{ij}) = c_{k(ij)}. \tag{4.5}$$

The $N_g \times N_g$ natrix

$$\mathbf{C}_k = \{c_{k(ij)}\} \tag{4.6}$$

gives all the segregation probabilities for the gamete γ_k in the population $\boldsymbol{\Gamma}$, and the vector

$$\mathbf{c}_{(ij)} = (c_{1(ij)}, c_{2(ij)}, \dots, c_{N_g(ij)}) \tag{4.7}$$

with $\sum_{k=1}^{N_g} c_{k(ij)} = 1$ represents the *segregation distribution* of the gametes $(\gamma_1, \gamma_2, \ldots, \gamma_{N_g})$ for the genotype Γ_{ij}.

If the genotype distribution matrix

$$\mathbf{Z} = \{z_{ij}\}. \tag{4.8}$$

corresponds to the genotype matrix $\mathbf{\Gamma} = \{\Gamma_{ij}\}$, then

$$\Pr\{\gamma_k, \Gamma_{ij}\} = c_{k(ij)} z_{ij}, \tag{4.9}$$

and the probability of the occurrence of a gamete γ_k in the population $\mathbf{\Gamma}$ is

$$\Pr(\gamma_k, \mathbf{\Gamma}) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} c_{k(ij)} z_{ij} = g_k^{(1)}, \tag{4.10}$$

for $k = 1, 2, \ldots, N_g$.

The vector

$$\mathbf{g}^{(1)} = (g_1^{(1)}, g_2^{(1)}, \ldots, g_{N_g}^{(1)}) \tag{4.11}$$

with $\sum_{k=1}^{N_g} g_k^{(1)} = 1$, represents the probabilities of occurrence of the gametes $(\gamma_1, \gamma_2, \ldots, \gamma_{N_g})$ in the population $\mathbf{\Gamma}$, or the *gamete probability vector* for the next generation. The $N_g \times N_g$ matrix

$$\mathbf{Z}^{(1)} = \mathbf{g}^{(1)}\mathbf{g}^{(1)'} \tag{4.12}$$

represents the genotype distribution in the next generation.

Assuming the segregation probabilities, $c_{k(ij)}$, in each generation are the same and applying (4.10) and (4.12), we obtain the following recurrence formula for the occurrence of a gamete γ_k in the n-th generation (gamete probability for the $(n+1)$-st generation)

$$\mathbf{g}_k^{(n+1)} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} g_i^{(n)} g_j^{(n)} c_{k(ij)}. \tag{4.13}$$

II. MORE THAN ONE LOCUS WITH MULTIPLE ALLELES AT EACH LOCUS IN 2m-PLOID CELLS

Suppose that there are r loci and to the locus a_h there are assigned s_h alleles: $a_{h1}, a_{h2}, \ldots, a_{hs_h} (h = 1, 2, \ldots, r)$. The general patterns for the genotype and geamte constitution are given by (2.1) and (2.2) respectively. But now we have to distinguish between genotype in "broader" and "narrower" sense, so the cases of independent and linked loci should be treated separately.

5. Independent loci.

5.1. Number of different genotypes and gametes. Let $N_{z_1}, N_{z_2}, \ldots, N_{z_r}$ denote the numbers of possible genotypes corresponding to the loci $a_1, a_2, \ldots, a_r$ respectively. The number of all possible genotypes with respect to r independent loci is

$$N_2 = \prod_{h=1}^{r} N_{z_h} = \prod_{h=1}^{r} \binom{s_h+2m-1}{2m}. \tag{5.1}$$

Replacing $2m$ by m we obtain the number of possible gametes.

The number of exactly q_1-genic types in a_1, q_2-genic types in $a_2, \ldots, q_r$-genic types in a_r is

$$N_{q_1 \ldots q_r\text{-genic}} = \prod_{h=1}^{r} \binom{s_h}{q_h}\binom{2m-1}{q_h-1}. \tag{5.2}$$

In a similar way we can find the number of different gametes which can be produced by a given genotype, and the number of different genotypes which could produce a given gamete.

5.2. Segregation distributions. Under the assumption of independent segregation of all r loci, the segregation probabilities $c_{k(ij)}$ with respect to all r loci are the products of segregation probabilities at each locus.

Denoting by $c_{k(ij)}(a_h)$ the segregation probability at locus a_h, we have

$$\Pr(\gamma_k | \Gamma_{ij}) = c_{k(ij)} = \prod_{h=1}^{r} c_{k(ij)}(a_h), \tag{5.3}$$

where the genotype Γ_{ij} and the gamete γ_k are each defined with regard to all r loci analogously as in section 4.1.

6. Linked loci.

6.1. When the r loci are linked; each chromosome can be made up in

$$S = s_1 \cdot s_2 \cdot \ldots \cdot s_r \tag{6.1}$$

ways. Therefore, the number of ordered boxes" is now S, and the number of different "segregating" genotypes (i.e. in "narrow sense") is

$$N_z = \binom{S+2m-1}{2m}. \tag{6.2}$$

(see also Geiringer [8]-[10]). Replacing $2m$ by m we obtain the number of gametes in "narrow sense". The numbers $N_{g|z}$ or $N_{z|g}$ can be calculated using the generating function (3.8) or (3.12) respectively, where s is replaced by S.

To find the number of segregating genotypes which arc *exactly* q_1--genic with respect to locus a_1, q_2-genic with respect to locus a_2, ... and q_r-genic with respect to locus a_s, i.e. $N_{q_1 \dots q_s}$-genic, we use differences of the function

$$\Phi(x_1, \dots, x_r) = \binom{x_1 \dots x_r + 2m - 1}{2m}. \tag{6.3}$$

It can be shown that

$$N_{q_1 \dots q_s}\text{-genic} = \Big[\prod_{h=1}^{s} \binom{s_h}{q_h} \Delta_{x_h}^{q_h}\Big] \Phi(x_1, \dots x_r)|_{x=0,} \tag{6.4}$$

(1)

where the difference operator Δ_{x_h} is defined by

$$\Delta_{x_h} \Phi(x_1, \dots x_r) \tag{6.5}$$

$$= \Phi(x_1, \dots, x_{h-1}, x_h + 1, x_{h+1}, \dots, x_r) - \Phi(x_1, \dots, x_{h-1}, x_h, x_{h+1}, \dots, x_r).$$

The cases of $r = 2$ and $r = 3$ are discussed in more detail by Fisher [3].

EXAMPLE 5. Let $r = 2, s_1 = 5, s_2 = 3, 2m = 4, q_1 = 3, q_2 = 2$.

$$\Phi(x, y) = \binom{xy+3}{4},$$

$$\Delta_x^3 \Phi(x, y)|_{x=0} = \Phi(3, y) - \binom{3}{2} \Phi(2, y) + \binom{3}{1} \Phi(1, y) - \Phi(0, y)$$

$$= \binom{3y+3}{4} - 3\binom{2y+3}{4} + 3\binom{y+3}{4},$$

$$\Delta_y^2[\Delta_x^3 \Phi(x, y)|_{x=0}]|_{y=0} = \Big[\binom{3 \cdot 2 + 3}{4} - 3\binom{2 \cdot 2 + 3}{4} + 3\binom{2+3}{4}\Big] -$$

$$- \binom{2}{1}\Big[\binom{3 \cdot 1 + 3}{4} - 3\binom{2 \cdot 1 + 3}{4} + 3\binom{1+3}{4}\Big] = 30.$$

Therefore

$$N_{3;2} = \binom{5}{3}\binom{3}{2} \times 30 = 900.$$

6.2. Segregation distributions. Assuming as in the case of one locus, the gametic output $\boldsymbol{\gamma}' = (\gamma_1, \gamma_2, \dots, \gamma_{N_g})$, where each gamete γ_k is a composite gamete in respect to r (linked) loci and defined in "narrower sense", we obtain the population matrix $\boldsymbol{\Gamma}$ analogously as in the case of one locus.

The general considerations and formulae of section 4.1 are still valid, but a new problem arises in evaluation of the segregation probabilities, $c_{k(ij)}$. In the case of linkage we cannot simply take the products of $c_{k(ij)}(a_h)$ as in 5.2. The recombination fractions for all cross-overs should be taken

---

(1) Note that $\Delta_x^m = f(x+m) - \binom{m}{1} f(x+m-1) + \binom{m}{2} f(x+m-2) + (-1)^m f(x)$.

into account and also all possible modes of gamete formation. The mechanism of crossing-over and gamete formation in polysomic inheritance is not known precisely and so far the frequencies of occurrence of genotypes and gametes have only been estimated empirically.

References

[1] N. T. J. Bailey, *Introduction to the mathematical theory of genetic linkage*, Clarendon Press, Oxford 1961.

[2] W. Feller, *An introduction to probability theory and its applications*, New York 1950.

[3] R. A. Fisher, *The theory of linkage in polysomic inheritance*, Phil. Trans. Roy. Soc. B 233 (1947), pp. 55-87.

[4] —, *A class of enumerations of importance in genetics*, Proc. Roy. Soc. B 136 (1950), pp. 509-520.

[5] H. Geiringer, *On the probability theory of linkage in Mendelian heredity*, Ann. Math. Statist. 15 (1944), pp. 25-57.

[6] —, *Further remarks on linkage in Mendelian heredity*, Ann. Math. Statist. 16 (1945), pp. 390-393.

[7] —, *Contribution to the heredity theory of multivalents*, J. Math. Phys. 26 (1948), pp. 246-278.

[8] —, *Contribution to the linkage theory of autopolyploids*, I, Bull. Math. Biophys. 2 (1949), pp. 59-82.

[9] —, *Contribution to the linkage theory of autopolyploids*, II, Bull. Math. Biophys. 2 (1949), pp. 197-219.

[10] —, *On some mathematical problems arising in the development of Mendelian genetics*, J. Amer. Statist. Assoc. 44 (1949), pp. 526-547.

[11] M. P. A. Macmahon, *Combinatory analysis*, vols. I, II, Cambridge University Press, Cambridge, Mass., 1915.

[12] P. A. P. Moran, *The statistical processes of evolutionary theory*, Clarendon Press, Oxford 1962.

[13] H. J. Ryser, *Combinatorial mathematics*, Math. Assoc. of America, Rahay 1963.

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, U.S.A.

Received on 25. 8. 1967