W. KLONECKI (Poznań)

ON THE PARTITION OF THE GENOTYPIC FUNCTION*

1. Introduction. Although the study of consequences of the Mendelian Law in a randomly mating population was started by Pearson [4] in 1903, it was Fisher [2] who in 1918 laid the theoretical foundations of the subject. Inspired by this paper, Kempthorne deduced an important formula, published in his book [3]. This formula represents the partition of the genotypic function into orthogonal components. Subsequently, Kempthorne himself and several other authors used this formula to deduce further results which are important in the interpretation of experimental data concerned with population genetics.

The present paper is concerned with Kempthorne's formula. His original deduction is based on an ingenious symbolic partition of the genotype itself. However, the derivation and use of the formula are a little complex so, in view of this formula's importance in modern research, an effort to obtain it through an alternative procedure seems justified. The method used in the present paper is based on the consideration of the genotypic function itself, and of its conditional expectations in relation to various fields. Thus, the result obtained does not coincide with Kempthorne's formula, but is equivalent to it. It is my impression that this method throws a new light on Kempthorne's theorem and also that it simplifies the application.

2. The problem. Consider a randomly mating population of individuals and an inheritable characteristic T to be described as the genotypic function. We assume that the characteristic T is affected by genes in n different loci, each carried in a different chromosome, so that there is no linkage. We denote by n_a the number of different genes capable of occupying the a-th locus, $a = 1, 2, \ldots, n$. The s-th gene appropriate to the a-th locus will be denoted by A_s^a and its relative frequency in the population by p_s^a, with $\sum_{s=1}^{n_a} p_s^a = 1$.

In this setup, the genotypic array is

$$(1) \qquad \prod_{a=1}^{n} \left(\sum_{s=1}^{n_a} A_s^a p_s^a \right)^2.$$

Incidentally, if the gene symbols A_s^a are treated as dummy arguments customarily used in studies of probability generating functions, then (1) represents the probability generating function of the various genotypes. In fact, the formal expansion of (1) represents a sum of terms in products of the type

$$(2) \qquad \prod_{a=1}^{n} A_{i_a}^a A_{j_a}^a,$$

where i_a and j_a are arbitrary numbers selected from $1, 2, \ldots, n_a$ symbolizing a particular genotype. The coefficient of (2) in the expansion of (1) represents the probability of encountering the given genotype in the population.

For any $a = 1, 2, \ldots, n$ the symbol

$$(3) \qquad \Omega^a = \{A_1^a, A_2^a, \ldots, A_{n_a}^a\}$$

will denote the set of all genes appropriate to the a-th locus. In addition to the symbol Ω^a just defined it will be convenient to consider two other symbols $\Omega^{a\cdot}$ and $\Omega^{\cdot a}$ denoting the same set as (3), but representing the n_a different possibilities of inheritance in the a-th locus from father and mother, respectively. Thus, the set of all genotypes in the population considered is the product set

$$(4) \qquad \Omega = \Omega^{1\cdot} \times \Omega^{\cdot 1} \times \Omega^{2\cdot} \times \Omega^{\cdot 2} \times \ldots \times \Omega^{n\cdot} \times \Omega^{\cdot n}.$$

The subject of study, the genotypic function, appears as a realvalued function f defined on Ω. After what was said about the expansion of (1) it must be clear that the unconditional expectation f^0 of f can be obtained by expanding (1) formally, replacing each product (2) by the corresponding value of f, that is, by

$$(5) \qquad f\left(\prod_{a=1}^{n} A_{i_a}^a A_{j_a}^a \right)$$

and by summing the results.

The total number of elements of the set Ω is $\prod_{a=1}^{n} n_a^2$. However, from the point of view of inheritance not all of these elements are distinct. Thus, for example, if any two distinct elements ω' and ω'' of Ω differ only by the order of the same two different genes in the a-th locus, say the genes $A_k^a A_i^a$ in ω' and $A_i^a A_k^a$ in ω'', with $i \neq k$, then the values of the

genotypic function at ω' and ω'' will be the same and the two genotypes must be considered identical. It is easy to see that the different elements of Ω correspond to $\prod_{a=1}^{n} n_a(n_a+1)/2^n$ different genotypes.

Kempthorne's formula can be written in the following form [3]:

$$
\begin{aligned}
\prod_{a=1}^{n} A_{i_a}^{a} A_{j_a}^{a} &= \prod_{a} \varrho^a \varrho^a + \sum_{a} v_{i_a}^{a} \varrho^a \prod_{a' \neq a} \varrho^{a'} \varrho^{a'} + \sum_{a} \varrho^a v_{j_a}^{a} \prod_{a' \neq a} \varrho^{a'} \varrho^{a'} + \\
&+ \sum_{a' \neq a} v_{i_a}^{a} v_{j_a}^{a} \prod \varrho^{a'} \varrho^{a'} + \sum_{a,a'} v_{i_a}^{a} \varrho^a v_{i_{a'}}^{a'} \varrho^{a'} \prod_{a''} \varrho^{a''} \varrho^{a''} + \\
&+ \sum_{a,a'} v_{i_a}^{a} \varrho^a \varrho^{a'} v_{j_{a'}}^{a'} \prod_{a''} \varrho^{a''} \varrho^{a''} + \sum_{a,a'} v_{j_a}^{a} \varrho^a v_{i_{a'}}^{a'} \varrho^{a'} \prod_{a''} \varrho^{a''} \varrho^{a''} + \\
&+ \sum_{a,a'} \varrho^a v_{j_a}^{a} \varrho^{a'} v_{j_{a'}}^{a'} \prod_{a''} \varrho^{a''} \varrho^{a''} + \sum_{a,a',a''} v_{i_a}^{a} \varrho^a v_{i_{a'}}^{a'} \varrho^{a'} v_{i_{a''}}^{a''} \varrho^{a''} \prod_{a'''} \varrho^{a'''} \varrho^{a'''} + \text{etc.},
\end{aligned}
$$

(6)

where

$$
\varrho^a = \sum_{s=1}^{n_a} p_s^a A_s^a,
$$

(7)

$$
v_{i_a}^{a} = A_{i_a}^{a} - \varrho^a,
$$

(8)

$$
v_{j_a}^{a} = A_{j_a}^{a} - \varrho^a
$$

(9)

and $a = 1, 2, \ldots, n$. This is a symbolic representation of a particular genotype.

In any particular term on the right side of (6) the possible values for a, a', a'', and so forth, are different. Each term on the right side has to be expanded by writing, for example, $\varrho^{a''}$ as $\sum_{s=1}^{n_{a''}} p_s^{a''} A_s^{a''}$ and $v_{i_{a'}}^{a'}$ as

$$
A_{i_{a'}}^{a'} - \sum_{s=1}^{n_{a'}} p_s^{a'} A_s^{a'}.
$$

In order to obtain the value of the genotypic function after performing the necessary multiplications in (6), the genotypic values have to be inserted in place of the genotypic symbols. The successive terms are: the population genotypic mean, the additive effects of genes $A_{i_a}^{a}$, the additive effects of genes $A_{j_a}^{a}$, the dominance deviations associated with genes $A_{i_a}^{a}$ and $A_{j_a}^{a}$, and so on. The terms in this sum are orthogonal. My aim is to give a representation of the genotypic function directly, without the use of formula (6).

3. Conditional expectation given a set of loci. For the sake of simplicity we shall be dealing not with the arbitrary genotypic function f itself, but with the deviation f^* of f from its expectation f^0. Thus for every genotype

$$
f = f^0 + f^*.
$$

(10)

The advantage gained by the use of the *centered genotypic function* f^*, is that its unconditional expectation is zero.

As already mentioned, the genotypic function is defined over a product set Ω of $2n$ components of which n are denoted by $\Omega^{a\cdot}$ and the other n by $\Omega^{\cdot a}$, with $a = 1, 2, \ldots, n$. Now fix an arbitrary integer $m, 1 \leqslant m \leqslant 2n$, and select out of the $2n$ component sets $\Omega^{a\cdot}$ and $\Omega^{\cdot a}$ an arbitrary combination of m sets, say the sets

$$(11) \qquad\qquad \Omega^{u_1}, \Omega^{u_2}, \ldots, \Omega^{u_m},$$

where each u_i denotes an integer between unity and n either preceded or followed by a dot. If u_i equals either $k\cdot$ or $\cdot k$ then the set Ω^{u_i} is composed of n_k genes A_s^k with $s = 1, 2, \ldots, n_k$. Let s_{u_i} be any of these values of s. Then the symbol $A_{s_{u_i}}^{u_i}$ will denote a gene appropriate to the k-th locus specified by the particular subscript s_{u_i}.

Denote by

$$(12) \qquad\qquad \Omega(A_{s_{u_1}}^{u_1}, A_{s_{u_2}}^{u_2}, \ldots, A_{s_{u_m}}^{u_m})$$

the set of all genotypes that include the genes specified in parentheses in the symbol (12). Thus, for example, $\Omega(A_1^{1\cdot})$ will denote the set of all the genotypes which include the gene $A_1^{1\cdot}$ in the "paternal" sublocus of the first of the loci considered. Similarly, $\Omega(A_1^{1\cdot}, A_2^{\cdot 1})$ will denote the set of all genotypes which include the genes $A_1^{1\cdot}$ and $A_2^{\cdot 1}$ in the "paternal" and "maternal" subloci of the first locus, etc.

Notice that if the subscripts $u_1, u_2, \ldots, u_m$ are kept constant, but the value of at least one of the subscripts $s_{u_1}, s_{u_2}, \ldots, s_{u_m}$ is changed, then the resulting set of genotypes of the form (12) will be disjoint from the original set (12). For example, if s_{u_1} is replaced by $s'_{u_1} \neq s_{u_1}$, then each of the genotypes in $\Omega(A_{s'_{u_1}}^{u_1}, \ldots, A_{s_{u_m}}^{u_m})$ will differ from every genotype of the set (12) by at least one gene, namely by the gene in the locus corresponding to the superscript u_1. On the other hand, two sets of the form (12) corresponding to two different systems of subscripts $u_1, u_2, \ldots, u_m$ need not be disjoint.

Denote by $G_{u_1, u_2, \ldots, u_m}$ the set of all sets (12) of genotypes corresponding to a fixed system of superscripts $u_1, u_2, \ldots, u_m$. These are sets (12) distinguished by different systems of subscripts $s_{u_1}, s_{u_2}, \ldots, s_{u_m}$. The union of all elements of $G_{u_1, u_2, \ldots, u_m}$ is the set Ω of all genotypes. The term conditional expectation of the centered genotypic function given the set of loci $u_1, u_2, \ldots, u_m$ or the conditional expectation with respect to $G_{u_1, u_2, \ldots, u_m}$ will be used to designate the function

$$(13) \qquad\qquad f_{u_1, u_2, \ldots, u_m} = E[f^* \mid G_{u_1, u_2, \ldots, u_m}]$$

defined on Ω as follows: on each set (12) the function (13) is constant and equal to the conditional expectation

$$(14) \qquad E[f^* \mid A^{u_1}_{s_{u_1}}, A^{u_2}_{s_{u_2}}, \ldots, A^{u_m}_{s_{u_m}}].$$

Notice that, if $m = 2n$, then each set of the form (12) is composed of just one element, namely of the genotype specified by the symbols in parentheses in (12). Also in this case the number of elements of $G_{1\cdot,\cdot1,2\cdot,\cdot2,\ldots,n\cdot,\cdot n}$ is the number of all the elements of Ω corresponding to $\prod_{a=1}^{n} n_a(n_a+1)/2^n$ different genotypes. Thus, if $m = 2n$, then

$$(15) \qquad f^* = E[f^* \mid G_{1\cdot,\cdot1,\ldots,n\cdot,\cdot n}].$$

The functions $f_{u_1,u_2,\ldots,u_m}$ have the following properties.

(i) They are defined over Ω for any number $m \leqslant 2n$ of subscripts u_i each of which may have a value $k\cdot$ or $\cdot k$, with $k = 1, 2, \ldots, n$. No two of these subscripts can have the same value, so that $u_i \neq u_j$ if $i \neq j$.

(ii) Because of the familiar property of conditional expectations,

$$(16) \qquad E[f_{u_1,\ldots,u_{m-1},u_m} \mid G_{u_1,\ldots,u_{m-1}}] = f_{u_1,u_2,\ldots,u_{m-1}},$$

and therefore for the unconditional expectation we have

$$(17) \qquad E[f_{u_1,u_2,\ldots,u_m}] = E[f^*] = 0.$$

(iii) For any two disjoint systems of positive integers $(u_1, u_2, \ldots, u_r)$ and $(w_1, w_2, \ldots, w_s)$, none exceeding $2n$, we have

$$(18) \qquad E[f_{u_1,u_2,\ldots,u_r} \mid G_{w_1,w_2,\ldots,w_s}] = E[f_{u_1,u_2,\ldots,u_r}] = 0.$$

This property results from the assumptions of no linkage between the n loci and of random mating. In fact, whatever system of genes in loci $w_1, w_2, \ldots, w_s$ we choose to fix, the relative frequencies of the various genes in loci $u_1, u_2, \ldots, u_r$ will not be affected. Hence, the conditional expectation on the left side of (18) must be equal to the unconditional expectation.

4. The partition of the genotypic function. The functions defined by (13) will be used to determine another class of functions, which, as we shall see, are orthogonal and add up to f^*.

DEFINITION. For m between limits, $1 \leqslant m \leqslant 2n$, let

$$(19) \qquad F_{u_1,u_2,\ldots,u_m} = f_{u_1,u_2,\ldots,u_m} - \sum_{t=1}^{m-1} \sum_{(s_1,s_2,\ldots,s_t)} F_{u_{s_1},u_{s_2},\ldots,u_{s_t}},$$

where the multiple summation extends over all systems of subscripts $(s_1, s_2, \ldots, s_t)$ such that $1 \leqslant s_1 < s_2 < \ldots < s_t \leqslant m$.

In particular for $m = 1, 2, 3$ we have

$$(20) \qquad F_{u_1} = f_{u_1},$$

$$(21) \qquad F_{u_1 u_2} = f_{u_1 u_2} - F_{u_1} - F_{u_2},$$

$$(22) \qquad F_{u_1 u_2 u_3} = f_{u_1 u_2 u_3} - F_{u_1} - F_{u_2} - F_{u_3} - F_{u_1 u_2} - F_{u_1 u_3} - F_{u_2 u_3}.$$

The functions $F_{u_1, u_2, \ldots, u_m}$ have the following properties.

(i) They are defined over Ω for any number $m \leqslant 2n$ of subscripts $u_1, u_2, \ldots, u_m$ each of which may have a value $k\cdot$ or $\cdot k$ with $k = 1, 2, \ldots, n$. No two of these subscripts can have the same value.

(ii) For any three disjoint systems of subscripts $(u_1, u_2, \ldots, u_p)$, $(v_1, v_2, \ldots, v_r)$, $(w_1, w_2, \ldots, w_s)$ none exceeding $2n$, we have

$$(23) \qquad E[F_{u_1, u_2, \ldots, u_p, v_1, v_2, \ldots, v_r} \mid G_{u_1, u_2, \ldots, u_p, w_1, w_2, \ldots, w_s}]$$
$$= E[F_{u_1, u_2, \ldots, u_p, v_1, v_2, \ldots, v_r} \mid G_{u_1, u_2, \ldots, u_p}].$$

If $p = 0$, then

$$(24) \qquad E[F_{v_1, v_2, \ldots, v_r} \mid G_{w_1, w_2, \ldots, w_s}] = E[F_{v_1, v_2, \ldots, v_r}] = 0.$$

(iii) If $1 \leqslant s < m \leqslant 2n$, then

$$(25) \qquad E[F_{u_1, u_2, \ldots, u_s, \ldots, u_m} \mid G_{u_1, u_2, \ldots, u_s}] = 0.$$

This property will be proved by induction. For $m = 2$ it follows from (16), (18) and (19), that

$$(26) \qquad E[F_{u_1, u_2} \mid G_{u_1}] = E[f_{u_1, u_2} - F_{u_1} - F_{u_2} \mid G_{u_1}]$$
$$= E[f_{u_1, u_2} \mid G_{u_1}] - E[f_{u_1} \mid G_{u_1}] - E[f_{u_2} \mid G_{u_1}]$$
$$= f_{u_1} - f_{u_1} - 0 = 0.$$

In order to prove (25) for m, provided that it holds for $m-1$, write (19) in the following form

$$(27) \qquad F_{u_1, u_2, \ldots, u_m} = f_{u_1, u_2, \ldots, u_m} - f_{u_1, u_2, \ldots, u_{m-1}} - R.$$

Notice that R in this formula is the sum of functions which have at most $m-1$ subscripts one of which is u_m. Bearing this in mind and using (23) and (25) for $m-1$, we obtain $E[R \mid G_{u_1, u_2, \ldots, u_s}] = 0$. The conditional expectation of $F_{u_1, u_2, \ldots, u_m}$ given $G_{u_1, u_2, \ldots, u_s}$ can now be computed as

$$(28) \qquad E[F_{u_1, u_2, \ldots, u_m} \mid G_{u_1, u_2, \ldots, u_s}] = E[f_{u_1, \ldots, u_m} - f_{u_1, \ldots, u_{m-1}} - R \mid G_{u_1, u_2, \ldots, u_s}]$$
$$= E[f_{u_1, u_2, \ldots, u_m} \mid G_{u_1, u_2, \ldots, u_s}] - E[f_{u_1, u_2, \ldots, u_{m-1}} \mid G_{u_1, \ldots, u_s}].$$

According to (16)

$$(29) \qquad E[f_{u_1, u_2, \ldots, u_m} \mid G_{u_1, u_2, \ldots, u_s}] = E[f_{u_1, u_2, \ldots, u_{m-1}} \mid G_{u_1, u_2, \ldots, u_s}] = f_{u_1, u_2, \ldots, u_s}.$$

Hence if $1 \leqslant s \leqslant m-1$

$$(30) \qquad E[F_{u_1, u_2, \ldots, u_m} \mid G_{u_1, u_2, \ldots, u_s}] = 0.$$

(iv) For two nonidentical systems of subscripts $(u_1, u_2, \ldots, u_m)$ and $(w_1, w_2, \ldots, w_r)$

$$(31) \qquad E[F_{u_1,u_2,\ldots,u_m} | F_{w_1,w_2,\ldots,w_r}] = 0.$$

This important property follows from (24). To prove it, let us first consider the case where the two systems of subscripts are disjoint. Then

$$(32) \qquad E[F_{u_1,u_2,\ldots,u_m} F_{w_1,w_2,\ldots,w_r}]$$

$$= \underset{(u_1,u_2,\ldots,u_m,w_1,w_2,\ldots,w_r)}{E} [F_{u_1,u_2,\ldots,u_m} F_{w_1,w_2,\ldots,w_r}]$$

$$= \underset{(u_1,u_2,\ldots,u_m)}{E} \left\{ F_{u_1,u_2,\ldots,u_m} \underset{(w_1,w_2,\ldots,w_r)}{E} [F_{w_1,w_2,\ldots,w_r}] \right\}.$$

Notice that

$$\underset{(w_1,w_2,\ldots,w_r)}{E} [F_{w_1,w_2,\ldots,w_r}] = E[F_{w_1,w_2,\ldots,w_r} | G_{(w_1,w_2,\ldots,w_r)^c}],$$

where $(w_1, w_2, \ldots, w_r)^c$ is the system of all subscripts that are not in $(w_1, w_2, \ldots, w_r)$. The property (24) implies that the conditional expectation given $G_{(w_1,w_2,\ldots,w_r)^c}$ (33) is identically equal to zero. Hence (31) follows. If the two systems of subscripts are not disjoint, then there is some s such that, say, $u_1 = w_1$, $u_2 = w_2$, \ldots, $u_s = w_s$, $s < r$. Hence,

$$E[F_{u_1,\ldots,u_m} F_{w_1,\ldots,w_r}] = E[F_{u_1,\ldots,u_m} F_{u_1,\ldots,u_s, w_{s+1},\ldots,w_r}]$$

$$\doteq \underset{(u_1,\ldots,u_m, w_{s+1},\ldots,w_r)}{E} [F_{u_1,\ldots,u_m} F_{u_1,\ldots,u_s, w_{s+1},\ldots,w_r}]$$

$$(34) \qquad = \underset{(u_1,\ldots,u_s)}{E} \left\{ \underset{(u_{s+1},\ldots,u_m, w_{s+1},\ldots,w_r)}{E} [F_{u_1,\ldots,u_m} F_{u_1,\ldots,u_s, w_{s+1},\ldots,w_r}] \right\}$$

$$= \underset{(u_1,\ldots,u_s)}{E} \left\{ \underset{(u_{s+1},\ldots,u_m)}{E} [F_{u_1,\ldots,u_m}] \underset{(w_{s+1},\ldots,w_r)}{E} [F_{u_1,\ldots,u_s, w_{s+1},\ldots,w_r}] \right\}.$$

Using the same argument as before we see that the terms in brackets are identically equal to zero, and therefore the functions $F_{u_1,u_2,\ldots,u_m}$ are orthogonal.

(v) The sum of all functions, $F_{u_1,u_2,\ldots,u_m}$, where $1 \leqslant m \leqslant 2n$ gives the function f^*,

$$(35) \qquad f^* = \sum_{a=1}^{n} F_{a\cdot} + \sum_{a=1}^{n} F_{\cdot a} + \sum_{a=1}^{n} F_{a\cdot,\cdot a} + \cdots + F_{1\cdot,\cdot 1, 2\cdot,\cdot 2,\ldots,n\cdot,\cdot n}.$$

In order to prove (35) put $m = 2n$ in (19),

$$(36) \qquad F_{u_1, u_2, \ldots, u_{2n}} = f_{u_1, u_2, \ldots, u_{2n}} - \sum_{t=1}^{2n-1} \sum_{(s_1, s_2, \ldots, s_t)} F_{u_{s_1}, u_{s_2}, \ldots, u_{s_t}},$$

where $1 \leqslant s_1 < s_2 < \ldots < s_t \leqslant 2n$.

As we have already mentioned, see (15), $f_{u_1, u_2, \ldots, u_{2n}} = f_{1 \cdot, \cdot 1, \ldots, n \cdot, \cdot n} = f^*$. Thus

$$(37) \qquad f^* = f_{u_1, u_2, \ldots, u_{2n}} = \sum_{t=1}^{2n} \sum_{(s_1, s_2, \ldots, s_t)} F_{u_{s_1}, u_{s_2}, \ldots, u_{s_t}},$$

where $1 \leqslant s_1 < s_2 < \ldots < s_t \leqslant 2n$, or

$$(38) \qquad f = f^0 + \sum_{a=1}^{n} F_{a \cdot} + \sum_{a=1}^{n} F_{\cdot a} + \sum_{a=1}^{n} F_{a \cdot, \cdot a} + \ldots + F_{1 \cdot, \cdot 1, 2 \cdot, \cdot 2, \ldots, n \cdot, \cdot n}.$$

This is Kempthorne's formula, representing the partition of the genotypic function. The values of the functions $F_{a \cdot}$, $F_{\cdot a}$, with $a = 1, 2, \ldots, n$ are the additive effects of genes, for example the additive effect of the gene A_s^a is equal to $F_{a \cdot}(A_s^a) = F_{\cdot a}(A_s^a)$. The other functions give the deviations associated with different groups of genes.

References

[1] E. R. Dempster, *Some genetic problems in controlled populations*, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 4 (1956), pp. 23-40.

[2] R. A. Fisher, *The correlation between relatives on the supposition of Mendelian inheritance*, Trans. Roy. Soc. Edin. 52 (1918), Part 2 (No. 15), pp. 399-433.

[3] O. Kempthorne, *An introduction to genetic statistics*, New York 1957.

[4] K. Pearson, *On a generalized theory of alternative inheritance with special reference to Mendel's Laws*, Phil. Trans. 203 (1903), A, pp. 53-87.

W. KLONECKI (Poznań)

O ROZKŁADZIE FUNKCJI GENOTYPOWEJ

STRESZCZENIE

Oryginalne wyprowadzenie znanej formuły Kempthorne'a o rozkładzie funkcji genotypowej na ortogonalne składniki opiera się na pewnym formalizmie. W tym wyprowadzeniu podstawowe jest pojęcie tzw. *szeregu genotypowego*. Autor przedstawia alternatywne wyprowadzenie tej formuły, opierające się na pojęciu funkcji ge-

notypowej i warunkowych oczekiwanych wartościach. Niech $\Omega^{\cdot i}$, $\Omega^{i \cdot}$ ($i = 1, 2, \ldots, n$) będą identycznymi zbiorami genów allelicznych odziedziczonych odpowiednio po ojcu i matce. Przestrzeń wszystkich genotypów ma wtedy postać

$$\Omega = \Omega^{1 \cdot} \times \Omega^{\cdot 1} \times \Omega^{2 \cdot} \times \Omega^{\cdot 2} \times \ldots \times \Omega^{n \cdot} \times \Omega^{\cdot n}.$$

Na przestrzeni tej określa się miarę produktową. Funkcja genotypowa f w takim ujęciu jest funkcją określoną na przestrzeni Ω. Oznaczmy przez Ef jej wartość oczekiwaną, a przez $f^* = f - Ef$. Niech $\Omega^{u_1}, \Omega^{u_2}, \ldots, \Omega^{u_m}$ — gdzie $1 \leqslant u_i \leqslant n$, a ponadto każde u_i zaopatrzone jest kropką albo z prawej, albo z lewej strony — będzie układem m różnych spośród rozważanych $2n$ zbiorów genów allelicznych. Przez $A^{u_i}_{s_{u_i}}$ oznaczymy gen należący do Ω^{u_i}, wyznaczony za pomocą wskaźnika s_{u_i}, a przez $\Omega(A^{u_1}_{s_{u_1}}, \ldots, A^{u_m}_{s_{u_m}})$ zbiór tych genotypów, które zawierają geny podane w nawiasie. Oznaczmy dalej przez $G_{u_1, u_2, \ldots, u_m}$ zbiór tych wszystkich zbiorów dla ustalonego układu wskaźników $(u_1, u_2, \ldots, u_m)$, a $f_{u_1, u_2, \ldots, u_m} = E[f^* \mid G_{u_1, u_2, \ldots, u_m}]$ niech będzie wartością oczekiwaną f^* pod warunkiem $G_{u_1, u_2, \ldots, u_m}$. Jeżeli określimy teraz indukcyjnie klasę funkcji wzorem

$$F_{u_1, u_2, \ldots, u_m} = f_{u_1, u_2, \ldots, u_m} - \sum_{t=1}^{m-1} \sum_{(s_1, \ldots, s_t)} F_{u_{s_1}, \ldots, u_{s_t}},$$

gdzie sumowanie rozciągnięte jest na wszystkie układy wskaźników $(s_1, \ldots, s_t)$ takich, że $1 \leqslant s_1 < \ldots < s_t \leqslant m$, to wspomniany wzór Kempthorne'a można napisać w postaci

$$f = Ef + \sum_{a=1}^{n} F_{a \cdot} + \sum_{a=1}^{n} F_{\cdot a} + \ldots + F_{1 \cdot, \cdot 1, 2 \cdot, \cdot 2, \ldots, n \cdot, \cdot n}.$$

W pracy dowodzi się, że składniki występujące po prawej stronie wzoru są ortogonalne.

В. КЛЬОНЕЦКИ (Познань)

О РАСПРЕДЕЛЕНИИ ГЕНОТИПНОЙ ФУНКЦИИ

РЕЗЮМЕ

Оригинальный вывод известной формулы Кемпторна о разложении генотипной функции на ортогональные слагаемые основан на некотором формализме. В этом выводе основным является понятие так называемого генотипного ряда. Автор даёт другой вывод этой формулы, основанный на понятии генотипной функции и на условных средних значениях. Пусть $\Omega^{\cdot i}$, $\Omega^{i \cdot}$ ($i = 1, 2, \ldots, n$) суть идентичные множества аллелических генов, унаследованных соответственно от отца и матери. Пространство всех генотипов имеет тогда вид

$$\Omega = \Omega^{1 \cdot} \times \Omega^{\cdot 1} \times \Omega^{2 \cdot} \times \Omega^{\cdot 2} \times \ldots \times \Omega^{n \cdot} \times \Omega^{\cdot n}.$$

В этом пространстве определяется мера как мера на декартовом произведении. При таком подходе генотипная функция f является функцией, определённой в пространстве Ω. Обозначим её среднее значение через Ef и пусть $f^* = f - Ef$. Пусть $\Omega^{u_1}, \Omega^{u_2}, \ldots, \Omega^{u_m}$, где $1 \leqslant u_i \leqslant n$, и, кроме того, u_i имеет точку с правой

либо с левой стороны, является системой m различных из $2n$ рассматриваемых множеств аллеличных генов. Через $A^{u_i}_{s_{u_i}}$ обозначим ген, принадлежащий к Ω^{u_i}, определённый индексом s_{u_i}, а через $\Omega(A^{u_1}_{s_{u_1}}, \ldots, A^{u_m}_{s_{u_m}})$ — множество тех генотипов, которые содержат члены, указанные в скобках. Обозначим через $G_{u_1,u_2,\ldots,u_m}$ множество всех этих множеств при фиксированной системе индексов $(u_1, u_2, \ldots, u_m)$ и пусть $f_{u_1,u_2,\ldots,u_m} = E[f^* | G_{u_1,u_2,\ldots,u_m}]$ будет средним значением f^* при условии $G_{u_1,u_2,\ldots,u_m}$. Если определить теперь по индукции класс функции формулой

$$F_{u_1,u_2,\ldots,u_m} = f_{u_1,u_2,\ldots,u_m} - \sum_{t=1}^{m-1} \sum_{(s_1,\ldots,s_t)} F_{u_{s_1},\ldots,u_{s_t}},$$

где суммирование распрастраняется на все системы индексов $(s_1, \ldots, s_t)$ таких, что $1 \leqslant s_1 < \ldots < s_t \leqslant m$, то упомянутая формула Кемпторпа может быть записана в виде

$$f = Ef + \sum_{a=1}^{n} F_{a\cdot} + \sum_{a=1}^{n} F_{\cdot a} + \ldots + F_{1\cdot,\cdot 1,2\cdot,\cdot 2,\ldots,n\cdot,\cdot n}.$$

В работе доказывается, что слагаемые, выступающие в правой стороне формулы, все ортогональны.