

102. Probability-theoretic Investigations on Inheritance. III₁. Further Discussions on Cross-Breeding.

By Yûsaku KOMATU.

Department of Mathematics, Tokyo Institute of Technology and
Department of Legal Medicine, Tokyo University.
(Comm. by T. FURUHATA, M.J.A., Oct. 12, 1951.)

1. Preliminaries.

In two previous papers¹⁾ we have developed a general theory of inheritance of a character consisting of m distinct genes denoted by A_i ($i = 1, \dots, m$), the inheritance of which is assumed to be subjected to Mendelian law. Especially, in §4 of I and §1 of II, we have promised to discuss the case in detail where the buffer effect grows gradually through several generations. In the present paper we shall treat such a problem. For the sake of brevity, we shall now confine ourselves to consider a population X composed of two sub-races X' and X'' . Suppose that these two sub-races are initially both *in equilibrium states*, then denoting the frequencies of genes A_i by p'_i and p''_i ($i = 1, \dots, m$), the frequencies of phenotypes are then given by the formulae

$$(1.1) \quad \begin{cases} A'_{ii} = p_i'^2, \\ \bar{A}'_{ij} = 2 p'_i p'_j; \end{cases} \quad \begin{cases} \bar{A}''_{ii} = p_i''^2, \\ \bar{A}''_{ij} = 2 p''_i p''_j \end{cases} \quad (i, j = 1, \dots, m; i < j).$$

Suppose further that two races X' and X'' are mixed at a rate $\lambda' : \lambda''$ ($\lambda' + \lambda'' = 1$), then the frequencies of the A_i in the limit distribution of X are, in view of the general result (1.7) of II, given by

$$(1.2) \quad p_i = \lambda' p'_i + \lambda'' p''_i \quad (i = 1, \dots, m),$$

and hence those of genotypes in the limit distribution of X are then, because of (1.8) of II expressed in the form

$$(1.3) \quad \bar{A}_{ii}^* = (\lambda' p'_i + \lambda'' p''_i)^2, \quad \bar{A}_{ij}^* = 2(\lambda' p'_i + \lambda'' p''_i)(\lambda' p'_j + \lambda'' p''_j) \quad (i \neq j).$$

On the other hand, let the initial distribution of X , i.e., the distribution of X at the moment of mixture, be denoted by $\bar{A}_{ij}(0)$ ($i \leq j$), while in the previous paper II it was denoted merely by \bar{A}_{ij} (cf. (1.9) of II). By general relations established in (1.9) of II, we then obtain

$$(1.4) \quad \begin{aligned} \bar{A}_{ii}(0) &= \lambda' \bar{A}'_{ii} + \lambda'' \bar{A}''_{ii} = \lambda' p_i'^2 + \lambda'' p_i''^2, \\ \bar{A}_{ij}(0) &= \lambda' \bar{A}'_{ij} + \lambda'' \bar{A}''_{ij} = 2(\lambda' p'_i p'_j + \lambda'' p''_i p''_j) \end{aligned} \quad (i, j = 1, \dots, m; i < j).$$

1) Y. Komatu, Probability-theoretic investigations on inheritance. I. Distribution of genes; II. Cross-breeding phenomena. Proc. Jap. Acad., **27** (1951), I, 371-377; II, 378-383, 384-387. These will be referred to as I and II respectively.

Before entering into the main discourse, we now make some preparations. Consider such matings alone each of which consists of the one chosen from a population with a distribution $\{p_i^{(1)}\}$ and the other that with a distribution $\{p_i^{(2)}\}$. The frequencies of genotypes in a group resulting from such matings are then given by corresponding terms of formal expansion of the product

$$\sum_{i=1}^m p_i^{(1)} \cdot \sum_{i=1}^m p_i^{(2)},$$

namely

$$(1.5) \quad \bar{A}_{ii} = p_i^{(1)} p_i^{(2)}, \quad \bar{A}_{ij} = p_i^{(1)} p_j^{(2)} + p_i^{(2)} p_j^{(1)} \\ (i, j = 1, \dots, m; i < j).$$

On the other hand, the frequencies of genes in the limit distribution of the group are, in view of (4.6) of I, given by

$$(1.6) \quad p_i = \bar{A}_{ii} + \frac{1}{2} \sum_{j \neq i} \bar{A}_{ij} = \frac{1}{2} (p_i^{(1)} + p_j^{(2)}) \quad (i = 1, \dots, m);$$

the same result would follow rather briefly by remembering that the present case corresponds to a particular one of (1.2) specialized as $\lambda' = \lambda'' = 1/2$.

As shown generally in (1.10) and (1.11) of II, we now have the relation between (1.3) and (1.4):

$$(1.7) \quad \bar{A}_{ii}(0) - \bar{A}_{ii}^* = \lambda' \lambda'' (p'_i - p''_i)^2, \\ \bar{A}_{ij}(0) - \bar{A}_{ji}^* = 2 \lambda' \lambda'' (p'_i - p''_i) (p'_j - p''_j) \quad (i, j = 1, \dots, m; i < j).$$

2. Preparatory considerations.

We now turn to our main discourse. We first investigate the frequencies of genotypes in the first generation when a partial cross-breeding takes place between X' and X'' . In such a cross-breeding process, all the matings are divided into three classes: the mating-class $X' \times X'$ between two individuals of X' , $X' \times X''$ between an individual of X' and that of X'' , and $X'' \times X''$ between two individuals of X'' . Three classes thus produced in the first generation will respectively be denoted by

$$(2.1) \quad X'^2 \equiv X' X', \quad X' X'', \quad X''^2 \equiv X'' X''.$$

Suppose now that X' and X'' have been mixed in the original generation at a rate $\lambda' : \lambda''$, as stated above. Let the total population be denoted by N ; then the numbers of individuals of X' and of X'' are, of course, equal to $\lambda' N$ and to $\lambda'' N$, respectively. Let further the cross-breeding $X' \times X''$ take place in $\mu_0^{(1)} N$ pairs; then the numbers of pairs of pure breedings $X' \times X'$ and $X'' \times X''$ are $(\lambda' N - \mu_0^{(1)} N)/2$ and $(\lambda'' N - \mu_0^{(1)} N)/2$, respectively. Hence, the matings of the classes $X' \times X'$, $X' \times X''$, $X'' \times X''$ then take place at the rate

$$(2.2) \quad (\lambda' - \mu_{0i}^{(1)}) : 2 \mu_{0i}^{(1)} : (\lambda'' - \mu_{0i}^{(1)}),$$

for which the identical relation holds: $(\lambda' - \mu_{0i}^{(1)}) + 2 \mu_{0i}^{(1)} + (\lambda'' - \mu_{0i}^{(1)}) = \lambda' + \lambda'' = 1$.

We first observe a homozygote A_{ii} . Its frequency in the first generation is contributed by each class, because of the first formula in (1.5), in amount,

$$(2.3) \quad X'X' : p_i'^2; \quad X'X'' : p_i'p_i''; \quad X''X'' : p_i''^2.$$

Hence, the total frequency of A_{ii} in the first generation is then, in view of (2.2), given by

$$(2.4) \quad \begin{aligned} \bar{A}_{ii}(1) &= (\lambda' - \mu_{0i}^{(1)})p_i'^2 + 2 \mu_{0i}^{(1)}p_i'p_i'' + (\lambda'' - \mu_{0i}^{(1)})p_i''^2 \\ &= \lambda'p_i'^2 + \lambda''p_i''^2 - \mu_{0i}^{(1)}(p_i' - p_i'')^2. \end{aligned}$$

Next, the frequency of a heterozygote A_{ij} ($i \neq j$) in the first generation is contributed by each class, because of the second formula in (1.5), in amount,

$$(2.5) \quad X'X' : 2 p_i'p_j'; \quad X'X'' : p_i'p_j'' + p_i''p_j'; \quad X''X'' : 2 p_i''p_j'',$$

whence its total frequency in the first generation is given by

$$(2.6) \quad \begin{aligned} \bar{A}_{ij}(1) &= (\lambda' - \mu_{0i}^{(1)})2 p_i'p_j' + 2 \mu_{0i}^{(1)}(p_i'p_j'' + p_i''p_j') + (\lambda'' - \mu_{0i}^{(1)})2 p_i''p_j'' \\ &= 2(\lambda'p_i'p_j' + \lambda''p_i''p_j'' - \mu_{0i}^{(1)}(p_i' - p_i'')(p_j' - p_j'')) \quad (i \neq j). \end{aligned}$$

If we now take the relations (1.4) into account, we may write the results (2.4) and (2.6) in the form

$$(2.7) \quad \begin{aligned} \bar{A}_{ii}(1) - \bar{A}_{ii}(0) &= -\mu_{0i}^{(1)}(p_i' - p_i'')^2, \\ \bar{A}_{ij}(1) - \bar{A}_{ij}(0) &= -2 \mu_{0i}^{(1)}(p_i' - p_i'')(p_j' - p_j'') \quad (i, j = 1, \dots, m; i < j). \end{aligned}$$

We thus conclude that, if only $p_i' = p_i''$ for some i , then, for every genotype partaken by the gene A_i , we get the relations

$$(2.8) \quad \bar{A}_{ij}(1) = \bar{A}_{ij}(0) \quad (j = 1, \dots, m),$$

regardless of the frequencies of remaining genes. Otherwise, i.e., if $p_i' \neq p_i''$ for some i , we obtain the inequality

$$(2.9) \quad \bar{A}_{ii}(1) < \bar{A}_{ii}(0);$$

in other words, the frequency of homozygote then diminishes in the next generation. With respect to the heterozygote A_{ij} ($i \neq j$), its frequency decreases or increases in the next generation according to the same or different signs of two factors $p_i' - p_i''$ and $p_j' - p_j''$.

From (1.7) and (2.7) we get

$$(2.10) \quad \begin{aligned} \bar{A}_{ii}(1) - \bar{A}_{ii}^* &= (\lambda' \lambda'' - \mu_{0i}^{(1)}) (p'_i - p''_i)^2, \\ \bar{A}_{ij}(1) - \bar{A}_{ij}^* &= 2 (\lambda' \lambda'' - \mu_{0i}^{(1)}) (p'_i - p''_i) (p'_j - p''_j) \end{aligned} \quad (i, j = 1, \dots, m; i < j).$$

If $p'_i = p''_i$ for some i , we have

$$(2.11) \quad \bar{A}_{ij}(1) = \bar{A}_{ij}^* \quad (j = 1, \dots, m),$$

while, if $p'_i \neq p''_i$ for some i , three cases are distinguished:

$$(2.12) \quad \lambda' \lambda'' > \mu_{0i}^{(1)}; \quad \lambda' \lambda'' < \mu_{0i}^{(1)}; \quad \lambda' \lambda'' = \mu_{0i}^{(1)}.$$

In the first case we get

$$(2.13) \quad \bar{A}_{ii}(1) > \bar{A}_{ii}^*$$

and further

$$(2.14) \quad \bar{A}_{ij}(1) > \bar{A}_{ij}^* \quad \text{or} \quad \bar{A}_{ij}(1) < \bar{A}_{ij}^* \quad (i < j)$$

according as $p'_i - p''_i$ and $p'_j - p''_j$ are of the same sign or not. In the second case of (2.12) we have only to invert the sense of all inequality signs in (2.13) and (2.14); in particular, the inequality $\bar{A}_{ii}(1) < \bar{A}_{ii}^*$ then appears, contrary to the case of random matings. In the last case of (2.12) we get always

$$(2.15) \quad \bar{A}_{ij}(1) = \bar{A}_{ij}^* \quad (i, j = 1, \dots, m),$$

regardless of the original distributions.

The limit distribution (1.3) of X is, of course, an equilibrium distribution. Hence, the relations (2.15) state that the distribution arrives at an equilibrium state soon in the first generation, provided that the last case of (2.12) occurs. But, this fact will also previously be expected as a matter of course.

At this occasion it is to be noticed that the frequencies of the gene A_i in three classes X^{12} , $X'X''$, X'^{12} are given by

$$(2.16) \quad p'_i, \quad \frac{1}{2}(p'_i + p''_i), \quad p''_i \quad (i = 1, \dots, m),$$

respectively, in which the result for $X'X''$ follows from (1.6).

We assume now that the birth-rates in classes of the original generation are equal each other. The rate of populations belonging to the classes X^{12} , $X'X''$, X'^{12} is then given by (2.2). Let us suppose that, in the first generation, the matings take place in portion $\mu_{01}^{(2)}$ between X^{12} and $X'X''$ being existent at the rates $\lambda' - \mu_{01}^{(1)}$ and $2\mu_{01}^{(1)}$ respectively, and in portion $\mu_{12}^{(2)}$ between $X'X''$ and X'^{12} being existent at the rates $2\mu_{01}^{(1)}$ and $\lambda'' - \mu_{01}^{(1)}$ respectively, and in portion $\mu_{01}^{(2)}$ between X^{12} and X'^{12} being existent at the rates $\lambda' - \mu_{01}^{(1)}$ and $\lambda'' - \mu_{01}^{(1)}$ respectively. Then, there are six sorts of possible mating-classes

$$(2.17) \quad \begin{aligned} X^{I^2} \times X^{I^2}, \quad X^{I^2} \times X^I X^{II}, \quad X^I X^{II} \times X^I X^{II}, \\ X^{I^2} \times X^{II^2}, \quad X^I X^{II} \times X^{II^2}, \quad X^{II^2} \times X^{II^2}. \end{aligned}$$

whose frequencies are given by the members of continued proportion

$$(2.18) \quad \begin{aligned} (\lambda' - \mu_{01}^{(1)} - \mu_{01}^{(2)} - \mu_{02}^{(2)}) : 2 \mu_{01}^{(2)} : (2 \mu_{01}^{(1)} - \mu_{01}^{(2)} - \mu_{12}^{(2)}) \\ : 2 \mu_{02}^{(2)} : 2 \mu_{12}^{(2)} : (\lambda'' - \mu_{01}^{(1)} - \mu_{12}^{(2)} - \mu_{02}^{(2)}), \end{aligned}$$

respectively. It will easily be seen that the sum of all terms contained in (2.18) is just equal to $\lambda' + \lambda'' = 1$. Frequencies of homozygote A_{ii} and of heterozygote A_{ij} ($i \neq j$), produced from each of the mating-classes in the second generation, are obtained, in view of (2.16), by the formulae (1.5). The results contained in (2.17) and (2.18) may be tabulated as follows.

class in the 2nd generation	mating-class in the 1st generation	frequency of each mating-class
X^{I^4}	$X^{I^2} \times X^{I^2}$	$\lambda' - \mu_{01}^{(1)} - \mu_{01}^{(2)} - \mu_{02}^{(2)}$
$X^{I^3} X^{II}$	$X^{I^2} \times X^I X^{II}$	$2 \mu_{01}^{(2)}$
$X^{I^2} X^{II^2}$	$\begin{cases} X^I X^{II} \times X^I X^{II} \\ X^{I^2} \times X^{II^2} \end{cases}$	$2 \mu_{01}^{(1)} - \mu_{01}^{(2)} - \mu_{12}^{(2)}$ $2 \mu_{02}^{(2)}$
$X^I X^{II^3}$	$X^I X^{II} \times X^{II^2}$	$2 \mu_{12}^{(2)}$
X^{II^4}	$X^{I^2} \times X^{II^2}$	$\lambda'' - \mu_{01}^{(1)} - \mu_{12}^{(2)} - \mu_{02}^{(2)}$

From the table we can calculate the frequencies of homozygote A_{ii} and of heterozygote A_{ij} ($i \neq j$) in the second generation, which will be denoted by $\bar{A}_{ii}(2)$ and $\bar{A}_{ij}(2)$ respectively. In fact, remembering that the frequencies of mating-classes are given in (2.18), we get

$$(2.19) \quad \begin{aligned} \bar{A}_{ii}(2) = & (\lambda' - \mu_{01}^{(1)} - \mu_{01}^{(2)} - \mu_{02}^{(2)}) p_i'^2 + \mu_{01}^{(2)} p_i'(p_i' + p_i'') \\ & + \frac{1}{4} (2 \mu_{01}^{(1)} - \mu_{01}^{(2)} - \mu_{12}^{(2)}) (p_i' + p_i'')^2 \\ & + 2 \mu_{02}^{(2)} p_i' p_i'' + \mu_{12}^{(2)} (p_i' + p_i'') p_i'' + (\lambda'' - \mu_{01}^{(1)} - \mu_{12}^{(2)} - \mu_{02}^{(2)}) p_i''^2, \end{aligned}$$

$$(2.20) \quad \begin{aligned} \bar{A}_{ij}(2) = & 2(\lambda' - \mu_{01}^{(1)} - \mu_{01}^{(2)} - \mu_{02}^{(2)}) p_i' p_j' + \mu_{01}^{(2)} (p_i'(p_j' + p_j'') + (p_i' + p_i'') p_j'') \\ & + \frac{1}{2} (2 \mu_{01}^{(1)} - \mu_{01}^{(2)} - \mu_{12}^{(2)}) (p_i' + p_i'') (p_j' + p_j'') + 2 \mu_{02}^{(2)} (p_i' p_j'' + p_i'' p_j') \\ & + \mu_{12}^{(2)} ((p_i' + p_i'') p_j'' + p_i'' (p_j' + p_j'')) + 2(\lambda'' - \mu_{01}^{(1)} - \mu_{12}^{(2)} - \mu_{02}^{(2)}) p_i'' p_j''. \end{aligned}$$

If we take the frequencies (2.4) and (2.6) in the first generation into account, we may write these formulae in the form

$$(2.21) \quad \begin{aligned} \bar{A}_{ii}(2) = & \bar{A}_{ii}(1) + \left\{ \frac{1}{2} \mu_{01}^{(1)} - \frac{1}{4} (\mu_{01}^{(2)} + \mu_{12}^{(2)} + 4 \mu_{02}^{(2)}) \right\} (p_i' - p_i'')^2, \\ \bar{A}_{ij}(2) = & \bar{A}_{ij}(1) + \left\{ \mu_{01}^{(1)} - \frac{1}{2} (\mu_{01}^{(2)} + \mu_{12}^{(2)} + 4 \mu_{02}^{(2)}) \right\} (p_i' - p_i'') (p_j' - p_j''). \end{aligned}$$

Comparing the last formulae further with (2.7), we thus deduce a result

$$(2.22) \quad \begin{aligned} \bar{A}_{ii}(2) - \bar{A}_{ii}(0) &= -\Gamma^{(2)}(p'_i - p''_i)^2, \\ \bar{A}_{ij}(2) - \bar{A}_{ij}(0) &= -2\Gamma^{(2)}(p'_i - p''_i)(p'_j - p''_j) \end{aligned} \quad (i, j = 1, \dots, m; i < j),$$

by introducing an abbreviation

$$(2.23) \quad \Gamma^{(2)} = \frac{1}{2}\mu_{01}^{(2)} + \frac{1}{4}(\mu_{01}^{(2)} + \mu_{12}^{(2)} + 4\mu_{02}^{(2)}).$$

In quite a similar manner as in (2.16), the frequencies of the gene A_i in six mating-classes (2.17) are given by

$$(2.24) \quad p'_i, \frac{1}{4}(3p'_i + p''_i), \frac{1}{2}(p'_i + p''_i), \frac{1}{2}(p'_i + p''_i), \frac{1}{4}(p'_i + 3p''_i), p''_i,$$

respectively. Since the frequencies of A_i in both classes $X'X'' \times X'X''$ and $X'^2 \times X''^2$ are identical, they may be unified into a single class. Such a unification reduces six classes in (2.17) to five which will be denoted by

$$(2.25) \quad X'^4, \quad X'^3X'', \quad X'^2X''^2, \quad X'X''^3, \quad X''^4.$$

These classes are regarded to appear, in the second generation, at the rate

$$(2.26) \quad \begin{aligned} (\lambda' - \mu_{01}^{(1)} - \mu_{01}^{(2)} - \mu_{02}^{(2)}) : 2\mu_{01}^{(2)} : (2\mu_{01}^{(1)} - \mu_{01}^{(2)} - \mu_{12}^{(2)} + 2\mu_{02}^{(2)}) \\ : 2\mu_{12}^{(2)} : (\lambda'' - \mu_{01}^{(1)} - \mu_{12}^{(2)} - \mu_{02}^{(2)}). \end{aligned}$$

In order to consider the passage from the second to the third generation, let five classes in (2.25) be numbered by 0, 1, 2, 3 and 4, respectively. Suppose now that the matings between the two classes with numbers u and v ($v > u$), i.e. $X'^{4-u}X''^{1u} \times X'^{4-v}X''^{1v}$, take place in the portion

$$(4.27) \quad \mu_{uv}^{(3)} \quad (u, v = 0, 1, \dots, 4; u < v).$$

Then, the similar considerations, the details of which will here be omitted, yield a result corresponding to (2.22) and (2.23); namely,

$$(2.28) \quad \begin{aligned} \bar{A}_{ii}(3) - \bar{A}_{ii}(0) &= -\Gamma^{(3)}(p'_i - p''_i)^2, \\ \bar{A}_{ij}(3) - \bar{A}_{ij}(0) &= -\Gamma^{(3)}(p'_i - p''_i)(p'_j - p''_j) \end{aligned} \quad (i, j = 1, \dots, m; i < j),$$

by introducing an abbreviation

$$(2.29) \quad \begin{aligned} \Gamma^{(3)} &= \frac{1}{2}\mu_{01}^{(1)} + \frac{1}{8}(\mu_{01}^{(2)} + \mu_{12}^{(2)} + 4\mu_{02}^{(2)}) \\ &+ \frac{1}{16}(\mu_{01}^{(3)} + \mu_{12}^{(3)} + \mu_{23}^{(3)} + \mu_{34}^{(3)} + 4(\mu_{02}^{(3)} + \mu_{13}^{(3)} + \mu_{24}^{(3)}) + 9(\mu_{03}^{(3)} + \mu_{14}^{(3)}) + 16\mu_{04}^{(3)}). \end{aligned}$$

Repeating the similar arguments, the formulae expressing the frequencies of genotypes in successive generations will be obtained step by step. However, such a stepwise progress will never come to a complete end. In order to finish the matter by one effort,

we attempt to seek inductively the general formulae for any generation. Thorough considerations on the above obtained results (2.7); (2.22), (2.23); (2.28), (2.29) will lead us to guess about that the frequencies of homozygote A_{ii} and of heterozygote A_{ij} ($i \neq j$) in the n th generation, which will be denoted by $\bar{A}_{ii}(n)$ and $\bar{A}_{ij}(n)$ respectively, may perhaps, given by the formulae

$$(2.30) \quad \begin{aligned} \bar{A}_{ii}(n) - \bar{A}_{ii}(0) &= -\Gamma^{(n)}(p'_i - p''_i)^2, \\ \bar{A}_{ij}(n) - \bar{A}_{ij}(0) &= -2\Gamma^{(n)}(p'_i - p'_j)(p'_j - p'_i) \end{aligned} \quad (i, j=1, \dots, m; i < j),$$

by introducing an abbreviation

$$(2.31) \quad \Gamma^{(n)} = \sum_{i=1}^{n-1} \frac{1}{2^{2i-1}} \sum_{k=1}^{2^{i-1}} k^2 \sum_{h=0}^{2^{i-1}-k} \mu_{h, h+k}^{(i)} + \frac{1}{2^{2n-2}} \sum_{k=1}^{2^{n-1}} k^2 \sum_{h=0}^{2^{n-1}-k} \mu_{h, h+k}^{(n)}.$$

Here the quantities

$$(2.32) \quad \mu_{uv}^{(i)} \quad (u, v = 0, 1, 2, 3, \dots, 2^{i-1}; u < v)$$

have similar meanings as those defined in (2.27) for case $l = 3$. Namely, let $2^{i-1} + 1$ classes existent in the $(i-1)$ th generation into which all the possible mating-classes in the $(i-2)$ th generation have been divided in conformity with the frequencies of genes, be denoted by

$$(2.33) \quad X^{i^{i-1}}, X^{i^{i-1}-1} X^{i^i}, X^{i^{i-1}-2} X^{i^{i^2}}, \dots, X^i X^{i^{i^2-1}-1}, X^{i^{i^2-1}}$$

and further, for the sake of brevity, be numbered by $0, 1, 2, \dots, 2^{i-1}-1, 2^{i-1}$, respectively. Then, it is supposed that the matings between the two classes with numbers u and v ($v > u$), i.e. $X^{i^{2^{i-1}-u}} X^{i^{i^u}} \times X^{i^{2^{i-1}-v}} X^{i^{i^v}}$, take place just in the portion given by $\mu_{uv}^{(i)}$.

The proof of the main results (2.30), (2.31) will be given in the next section.

—To be Continued—