104. Probability-theoretic Investigations on Inheritance.
III$_3$. Further Discussions on Cross-Breeding.
(Further Continuation.)

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.)

4. Consequences of main result.

In the last section we have proved the main result, stating

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

where

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

Having been established this result, we can attempt the similar considerations on the comparison between $\bar{A}_{ii}(n)$ or $\bar{A}_{ij}(n)$ and the corresponding quantities with respect to earlier generations, as done in a special case in § 2.

First, for instance, if $p'_i = p''_i$ for some fixed suffix i, we see that the relations

$$(4.1) \qquad \bar{A}_{ij}(n) = \bar{A}_{ij}(0) \qquad (j = 1,\dots, m\,;\ n = 1, 2,\dots)$$

are always valid, regardless of the frequencies of the remaining genes; the relations $\bar{A}_{ij}(n) = \bar{A}^*_{ij}$ $(j = 1,\dots, m)$ hold then also good. We thus conclude that, if the sub-races X' and X'' being composed into X possess an identical frequency of a gene A_i for some i, then every genotpye partaken by A_i remains constant with respect to generation, regardless of the frequencies of the remaining genes, in how manner the cross-breeding does take place. Here, it must, of course, be remembered that the condition of equilibrium states of both sub-races is kept to be assumed.

Naturally, in probability-theoretic treatment of general theory on inheritance phenomena, an immediate object of various considerations is the frequencies of the genes, and hence those of genotypes. Consequently, the immediately deducible concrete results concerning phenotypes refer only to phenotypes consisting of the genes dominant against no other genes, as O or M and N in ABO or MN blood types, respectively.

However, the above obtained result makes possible to deduce a general result also on phenotypes. In fact, as noticed above, if both sub-races X' and X'' being initially in equilibrium states possess an coincident frequency of a gene A_i, then the frequencies of the genotypes

$$A_{ij} \equiv A_i A_j \qquad (j = 1, \ldots, m) \tag{4.2}$$

in the composed population X remain constant with respect to generation, regardless of the frequencies of the remaining genes, in how manner the cross-breeding process does ever take place. Now, the phenotype represented by the gene A_i consists at any rate of some genotypes contained in (4.2). Hence, the frequency of the phenotype represented by A_i remains also constant with respect to generation, under the only assumption $p_i' = p_i''$ for that i.

Next, suppose that $p_i' \neq p_i''$ for some i. Since evidently the quantity $\Gamma^{(n)}$ is positive, we then have

$$A_{ii}(n) < \bar{A}_{ii}(0) \qquad (n = 1, 2, \ldots). \tag{4.3}$$

Indeed, if (and only if) all the $\mu_{uv}^{(1)}$ vanish, $\Gamma^{(n)}$ also becomes zero. But, this is the case where the cross-breeding process does not essentially take place and hence the both sub-races remain merely in a mosaic manner. It is therefore a matter of course that then $\bar{A}_{ii}(n) = \bar{A}_{ii}(0)$, i.e., the initial distribution remains constant. Such an extremely special case may be excluded from our considerations.

We thus conclude that *in case $p_i' \neq p_i''$ for some i the frequency of a homozygote A_{ii} in every subsequent generation will diminish compared with that in the initial generation, while in case $p_i' = p_i''$ it remains constant.* The frequency of a heterozygote $A_{ij} (i \neq j)$ in every subsequent generation will decrease or increase compared with that in the initial generation according as $p_i' - p_i''$ and $p_j' - p_j''$ are of the same sign or not. Only in case $p_i' = p_i''$ or $p_j' = p_j''$ it remains constant with respect to generation, as already shown in (4.1).

The comparison of the distribution in the nth generation with that in the limit distribution may be performed by means of the relations obtained from (1.7) and (2.30) stating

$$\left.\begin{aligned} \bar{A}_{ii}(n) - \bar{A}_{ii}^* &= (\lambda'\lambda'' - \Gamma^{(n)})(p_i' - p_i'')^2, \\ \bar{A}_{ij}(n) - \bar{A}_{ij}^* &= 2(\lambda'\lambda'' - \Gamma^{(n)})(p_i' - p_i'')(p_j' - p_j'') \end{aligned}\right\} \quad (i, j = 1, \ldots, m;\ i < j). \tag{4.4}$$

If $p_i' = p_i''$ for some i, the frequencies $\bar{A}_{ij}(n)$ coincide for every j also with the corresponding ones $\bar{A}_{ij}^*$ in the limit distribution. Hence, together with (4.1), we then have

$$\bar{A}_{ij}(n) = \bar{A}_{ij}(0) = \bar{A}_{ij}^* \quad (j = 1, \ldots, m;\ n = 1, 2, \ldots). \tag{4.5}$$

If $p_i' \neq p_i''$ for some i, three cases are distinguished:

$$\lambda'\lambda'' > \Gamma^{(n)}; \quad \lambda'\lambda'' < \Gamma^{(n)}; \quad \lambda'\lambda'' = \Gamma^{(n)}. \tag{4.6}$$

In the first case we get

$$(4.7)\qquad \bar{A}_{ii}(n) > \bar{A}_{ii}^{*} \qquad (i = 1, \dots, m;\ n = 1, 2, \dots),$$

and further

$$(4.8)\qquad \bar{A}_{ij}(n) > \bar{A}_{ij}^{*} \quad \text{or} \quad \bar{A}_{ij}(n) < \bar{A}_{ij}^{*} \quad (i < j)$$

according as $p_i' - p_i''$ and $p_j' - p_j''$ are of the same sign or not. In the second case of (4.6) we have only to invert the sense of all inequality signs in (4.7) and (4.8); in particular the inequality $\bar{A}_{ii}(n) < \bar{A}_{ii}^{*}$ then appears, contrary to the case of random matings. In the last case of (4.6) we get always

$$(4.9)\qquad \bar{A}_{ij}(n) = \bar{A}_{ij}^{*} \qquad (i, j = 1, \dots, m),$$

regardless of the original distributions; namely, the distribution in the nth generation becomes to an equilibrium distribution coinciding also with the limit distribution. Since the only possible equilibrium distribution in the composed population is its limit distribution, we may state the above mentioned fact also as follows:

A necessary and sufficient condition that an equilibrium distribution is realized in the nth generation is given by the validity of the last relation of (4.6), or fully described,

$$(4.10)\qquad \sum_{l=1}^{n-1} \frac{1}{2^{2l-1}} \sum_{k=1}^{2^{l-1}} k^2 \sum_{h=0}^{2^{l-1}-k} \mu_{h,h+k}^{(l)} + \frac{1}{2^{2n-2}} \sum_{k=1}^{2^{n-1}} k^2 \sum_{h=0}^{2^{n-1}-k} \mu_{h,h+k}^{(n)} = \lambda'\lambda''.$$

Now, the number of the whole possible classes in the $(l-1)$th generation is $2^{l-1}+1$, as listed in (2.33). Hence, the number of the independent quantities $\mu_{uv}^{(l)}$ $(u, v = 0, 1, 2, \dots, 2^{l-1};\ u < v)$ in (2.30), which provide the manner of cross-breedings at the passage from this to the next generation, is

$$(4.11)\qquad \binom{2^{l-1}+1}{2} = (2^{l-1}+1)2^{l-2}$$

Thus, the total number of the independent quantities

$$(4.12)\qquad \mu_{uv}^{(l)} \qquad (u, v = 0, 1, 2, \dots, 2^{l-1};\ u < v;\ l = 1, \dots, n),$$

providing the manner of cross-breedings up to the nth generation, is equal to the sum

$$(4.13)\qquad \sum_{l=1}^{n} (2^{l-1}+1)2^{l-2} = \frac{1}{3}(2^{n-1}+2)(2^{n}-1).$$

The quantity $\Gamma^{(n)}$ standing on the left-hand side of the relation (4.10) is a homogeneous linear expression with respect to the quantities (4.12). The relation (4.10), representing the condition that an equilibrium distribution is realized in the nth generation, is a unique equation between these quantities whose total number is equal to (4.13). Since the last number increases very rapidly with n, there remains still an extraordinarily large degree of freedom between the quantities (4.12) satisfying the unique equation (4.10). Thus, there exist very various manners of cross-breedings which result an equilibrium state in the nth generation.

All the quantities (4.12) must naturally satisfy the inequalities

$$0 \leqq \mu_{uv}^{(l)} \leqq 1/2. \tag{4.14}$$

On the other hand, the frequencies of all the mating-classes in the nth generation, listed in a table of the last section, must, of course, lie in the closed interval $0 \leqq x \leqq 1$, and hence the further conditions

$$\begin{gathered}
\sum_{l=1}^{n} \sum_{k=1}^{2^{l-1}} \mu_{0k}^{(l)} \leqq \lambda', \qquad \sum_{l=1}^{n} \sum_{k=0}^{2^{l-1}-1} \mu_{2^{l}-1,k}^{(l)} \leqq \lambda'', \\
0 \leqq \sum_{l=n+1-d(\alpha)}^{n} \Big(2 \sum_{0 \leqq h < 2^{l-n-1}\alpha} \mu_{h,2^{l-n}\alpha-h}^{(l-1)} - \sum_{k=0}^{2^{l-1}} \mu_{2^{l-n}\alpha,k}^{(l)} \Big) \leqq 1 \\
\hfill (1 \leqq \alpha < 2^{n-2}), \\
0 \leqq \sum_{l=n+1-d(\alpha)}^{n} \Big(2 \sum_{2^{l-2} \geqq h > 2^{l-n-1}\alpha} \mu_{h,2^{l-n}\alpha-h}^{(l-1)} - \sum_{k=0}^{2^{l-1}} \mu_{2^{l-n}\alpha,h}^{(l)} \Big) \leqq 1 \\
\hfill (2^{n-2} \leqq \alpha < 2^{n-1})
\end{gathered} \tag{4.15}$$

must also be satisfied; in fact, the total sum of frequencies of all the mating-classes each of which is a non-negative number, is equal to $\lambda' + \lambda'' = 1$, as noticed explicitly at the end of the preceding section.

Conversely, if the relations (4.14) and (4.15) are all satisfied, then a realizable manner of cross-breedings can be provided, provided that the population X contains sufficiently large number of individuals such that the fractions of an individual may be cut off.

Now the quantities (4.12) provide not only a manner of cross-breedings, as stated frequently, but also its degree qualitatively. In practice, starting from the initial generation when the new population X is composed, the cross-breeding process will extend over gradually. Then, at the beginning stage of the process all the quantities (4.12) may be expected very small compared with $\lambda'\lambda''$, and hence in earlier generations the first inequalities of (4.6) will surely remain valid, and consequently also (4.7). But, if the cross-breeding process proceeds further, it may be possible that the first case in (4.6) changes into the second. However, the second case in (4.6) represents, so to speak, a state of supermixture. If such a state would be reached temporarily for a while, it will show a tendency toward a perfect mixture represented by the last relation in (4.6), provided that in practice the buffer effect acts.

It will be a very interesting problem, in particular from practical point of view, to consider the case, where the process of cross-breeding proceeds very slowly. In such a case, the quantity $\Gamma^{(n)}$ will increase monotonously with n through several early generations so that the corresponding distributions will tend to the limit equilibrium distribution step by step. Hence, it will be expected that, based on such a tendency, the frequency of every homozygote decreases monotonously. This remarkable fact will be able to illustrated by such a phenotype which consists of only one genotype, for instance, O; M, N; q in ABO; MN; Q blood type, respectively.

We now compare the distribution in the nth generation with that in the preceding one. We obtain, immediately from our main formulae, the relations

$$\begin{aligned} &\bar{A}_{ii}(n)-\bar{A}_{ii}(n-1) = (\Gamma^{(n-1)}-\Gamma^{(n)})\,(p'_i-p''_i)^2, \\ &\qquad\qquad\qquad\qquad (i,\, j = 1, \ldots, m\,;\ i<j), \\ (4.16)\quad &\bar{A}_{ij}(n)-\bar{A}_{ij}(n-1) = 2(\Gamma^{(n-1)}-\Gamma^{(n)})(p'_i-p''_i)\,(p'_j-p''_j) \end{aligned}$$

where the factor standing in the right-hand side is, in view of (3.16), given by

$$(4.17)\quad \Gamma^{(n-1)}-\Gamma^{(n)} = \frac{1}{2^{2n-2}}\left(2\sum_{k=1}^{2^{n-1}} k^2 \sum_{h=0}^{2^{n-2}-k} \mu^{(n-1)}_{h,h+k} - \sum_{k=1}^{2^{n-1}} k^2 \sum_{h=0}^{2^{n-1}-k} \mu^{(n)}_{h,h+k}\right).$$

We thus conclude that the frequency of every homozygote in the nth generation is greater or less than that in the $(n-1)$th one according that the difference $\Gamma^{(n-1)}-\Gamma^{(n)}$ is positive or negative. It will be worth mentioning that, in particular, the frequencies of all homozygtes vary in the same direction, at any passage to the next generation, depending only on the sign of the difference (4.17). Consequently, for instance, in MN blood type we see that the both phenotypes M and N will increase or decrease simultaneously. The distribution in the nth generation coincides with that in the $(n-1)$th one if and only if $\Gamma^{(n-1)} = \Gamma^{(n)}$, i.e.,

$$(4.18)\qquad 2\sum_{k=1}^{2^{n-2}} k^2 \sum_{h=1}^{2^{n-2}-k} \mu^{(n-1)}_{h,h+k} = \sum_{k=1}^{2^{n-1}} k^2 \sum_{h=0}^{2^{n-1}-k} \mu^{(n)}_{h,h+k}\,,$$

a relation containing merely $\mu^{(l)}_{uv}$ with $l = n-1$ and $l = n$.

We shall close the present section by considering some special cases. Suppose first that

$$(4.19)\qquad \mu^{(l)}_{uv} = 0 \qquad (u,\, v = 0, 1, 2, \ldots, 2^{n-1}).$$

This is the case where the matings in the $(n-1)$th generation take place separately in each of the $(2^{n-1}+1)$ classes. Among the possible classes in the nth generation, the frequencies of the classes

$X'^{2n-(2\alpha-1)}X''^{(2\alpha-1)}$ $(1 \leqq \alpha \leqq 2^{n-1})$ then all vanish, while those of the classes $X'^{2n-2\alpha}X''^{2\alpha}$ $(1 \leqq \alpha < 2^{n-1})$ are equal to the frequencies of $X'^{2^{n-1}-\alpha}X''^{\alpha}$ $(0 \leqq \alpha \leqq 2^{n-1})$, respectively. In fact, for example, the frequency of a class $X'^{2^{n-1}-\alpha}X''^{\alpha}$ with $2 \leqq \alpha \leqq 2^{n-2}$ is, under the assumption (4.19), given by

$$\sum_{l=n+1-d(\alpha)}^{n}\Big(2\sum_{0\leqq h<2^{l-n-1}\alpha}\mu^{(l-1)}_{h,\,2^{l-n}\alpha-h}-\sum_{k=0}^{2^{l-1}}\mu^{(l)}_{2^{l-n}\alpha,\,k}\Big)$$

$$=\sum_{l=n-d(\alpha/2)}^{n}\Big(2\sum_{0\leqq h<2^{l-n}\alpha/2}\mu^{(l-1)}_{h,\,2^{l-n+1}\alpha/2-h}-\sum_{k=0}^{2^{l-1}}\mu^{(l)}_{2^{l-n+1}\alpha/2,\,k}\Big)$$

$$=2\sum_{0\leqq h<\alpha/2}\mu^{(n-1)}_{h,\,\alpha/2-h}+\sum_{l=n-d(\alpha/2)}^{n-1}\Big(2\sum_{0\leqq h<2^{l-n}\alpha/2}\mu^{(l-1)}_{h,\,2^{l-n+1}\alpha/2-h}-\sum_{k=0}^{2^{l-1}}\mu^{(l)}_{2^{l-n+1}\alpha/2,\,k}\Big)$$

which coincides just with the frequency of $X'^{2^{n-1}-\alpha}X''^{\alpha}$. The same result follows also for remaining α. However, such a fact is also an immediately evident consequence of the assumption (4.19). Under the same assumption (4.19), it would seem that the whole distribution in the nth generation would coincide with that in the $(n-1)$th generation. But, we can deduce only an inequality

$$(4.20)\qquad \Gamma^{(n-1)}-\Gamma^{(n)}=\frac{1}{2^{2n-3}}\sum_{k=1}^{2^{n-2}}k^2\sum_{h=0}^{2^{n-2}-k}\mu^{(n-1)}_{h,\,h+k}\geqq 0\,,$$

which implies that $\Gamma^{(n-1)}-\Gamma^{(n)}$ is strictly positive when even a little cross-breeding has taken place till $(n-1)$th generation. We then have

$$(4.21)\qquad \lambda'\lambda''-\Gamma^{(n)}>\lambda'\lambda''-\Gamma^{(n-1)}\,,$$

in view of which we conclude, based upon (4.4), that the distribution in the nth generation deviates more distant from the equilibrium state than that in the $(n-1)$th generation. It will be a reason of such a phenomenon that the diminution of the homozygotes caused by matings between different classes is suppressed.

If, in addition to (4.19), we suppose further that the conditions

$$(4.22)\qquad \mu^{(n-1)}_{uv}=0 \qquad (u,v=0,1,2,\ldots\ldots,2^{n-2})$$

are satisfied, i.e., if, at passage from the $(n-1)$th generation to nth as well as from $(n-2)$th to $(n-1)$th, no mating takes place between different classes, then the equality sign holds in the inequality (4.20), whence it follows that the distributions of the $(n-1)$th and of the nth generation coincide each other.

—To be Continued—