85. Probability theoretic Investigations on Inheritance. II₁. Cross-Breeding Phenomena.

By Yûsaku Komatu.

Department of Mathematics, Tokyo Institute of Technology and
Department of Legal Medicine, Tokyo University.

(Comm. by T. Furuhata, m.j.a., July 12, 1951.)

1. Limit distribution of composed population.

In a population, the distribution of an inherited character will arrive at an equilibrium state soon in the next generation, if matings take place at random with respect to the character, even when the distribution in initial generation is not in an equilibrium state. This is a fact that we have already noticed at the end of the previous paper[1]. In practice, however, particularly in such a case where the population cover a very wide range, it will not be expected that such random matings take place for once only. We should rather expect that this buffer effect grows gradually through many generations. Although we shall postpone the detailed discussion of this problem still later, we notice here a remarkable fact in a cross-breeding process.

As in the last section of I, we consider g races $X^{(\nu)}$ each of which possesses an equilibrium distribution. Denoting by

$$(1.1) \qquad p_i^{(\nu)} \qquad (i = 1, \ldots, m)$$

the frequencies of genes A_i in the race $X^{(\nu)}$, then the frequencies of genotypes are given by

$$(1.2) \qquad \bar{A}_{ii}^{(\nu)} = p_i^{(\nu)2}, \qquad \bar{A}_{ij}^{(\nu)} = 2p_i^{(\nu)}p_j^{(\nu)} \quad (i \neq j).$$

These quantities satisfy of course the fundamental relations

$$\sum_{i=1}^{m} p_i^{(\nu)} = 1, \quad \sum_{i \leq j} \bar{A}_i^{(\nu)} = 1 \qquad (\nu = 1, \ldots, g)$$

and moreover, as shown in (2.9) of I, we have

$$(1.3) \quad p_i^{(\nu)} = \bar{A}_{ii}^{(\nu)} + \frac{1}{2}\sum_{j \neq i} A_{ij}^{(\nu)} \quad (i = 1, \ldots, m; \nu = 1, \ldots, g).$$

Suppose now, as in §4 of I, that these races $X^{(\nu)}$ ($\nu = 1, \ldots, g$) are mixed at the rate $\lambda^{(\nu)}(> 0)$ with

$$(1.4) \qquad \sum_{\nu=1}^{g} \lambda^{(\nu)} = 1$$

the distribution of resulting population X will then be given by

$$(1.5) \qquad \bar{A}_{ii} = \sum_{\nu=1}^{g} \lambda^{(\nu)} \bar{A}_{ii}^{(\nu)}, \quad \bar{A}_{ij} = \sum_{\nu=1}^{g} \lambda^{(\nu)} \bar{A}_{ij}^{(\nu)}.$$

---

1) Y. Komatu, Probability-theoretic investigations on inheritance. I. Distribution of genes. Proc. Jap. Acad., 27 (1951), 371–377. This will be referred to as I.

Now, as already shown at the beginning of §4 of I, the distribution of this composed race cannot be in equilibrium state provided that all the component races have not an identical distribution. Under these circumstances we take as frequency p_i of each gene A_i in the race X, the frequency of A_i which will appear in its *limit distribution*, that is, an equilibrium distribution arrived at ultimately after buffer effect. In view of the result in §4 of I we have

$$(1.6) \qquad p_i = \bar{A}_{ii} + \tfrac{1}{2}\sum_{j \neq i} \bar{A}_{ij} \qquad (i = 1, \ldots, m).$$

Substituting the values (1.5) in the right-hand side of (1.3), we get

$$p_i = \sum_{\nu=1}^{g} \lambda^{(\nu)} \bar{A}_{ii}^{(\nu)} + \tfrac{1}{2}\sum_{j \neq i}\sum_{\nu=1}^{g} \lambda^{(\nu)} A_{ij}^{(\nu)} = \sum_{\nu=1}^{g} \lambda^{(\nu)}\left(\bar{A}_{ii}^{(\nu)} + \tfrac{1}{2}\sum_{j \neq i} \bar{A}_{ij}^{(\nu)} \right),$$

which may, in view of (1.3), be also written as

$$(1.7) \qquad p_i = \sum_{\nu=1}^{g} \lambda^{(\nu)} p_i^{(\nu)} \qquad (i = 1, \ldots, m).$$

These are evident relations which may also be previously expected.

In order to distinguish the frequency of genotypes $A_{ij}(i \leq j)$ in the limit distribution from that in the original one, we denote the former by $\bar{A}_{ij}^{*}$ while the latter is denoted by $\bar{A}_{ij}$. Because of (1.7) we obtain

$$(1.8) \qquad \begin{aligned} \bar{A}_{ii}^{*} &= p_i^2 = \left(\sum_{\nu=1}^{g} \lambda^{(\nu)} p_i^{(\nu)}\right)^2, \\ \bar{A}_i^{*} &= 2p_ip_j = 2\sum_{\nu=1}^{g} \lambda^{(\nu)} p_i^{(\nu)} \sum_{\nu=1}^{g} \lambda^{(\nu)} p_j^{(\nu)} \qquad (i \neq j). \end{aligned}$$

On the other hand, in view of (1.2) and (1.5), the original distribution is given by

$$(1.9) \qquad \bar{A}_{ii} = \sum_{\nu=1}^{g} \lambda^{(\nu)} p_i^{(\nu)2}, \qquad \bar{A}_{ij} = 2\sum_{\nu=1}^{g} \lambda^{(\nu)} p_i^{(\nu)} p_{\nu}^{(j)} \quad (i \neq j).$$

We now compare the original distribution of the race X with its limit distribution. Remembering the identical relation (1.4), we have, from (1.8) and (1.9), for difference between frequencies of homozygote A_{ii}:

$$(1.10) \quad \bar{A}_{ii} - \bar{A}_{ii}^{*} = \sum_{\nu=1}^{g} \lambda^{(\nu)} \sum_{\nu=1}^{g} \lambda^{(\nu)} p_i^{(\nu)2} - \left(\sum_{\nu=1}^{g} \lambda^{(\nu)} p_i^{(\nu)}\right)^2 = \sum_{<\nu} \lambda^{(\mu)} \lambda^{(\nu)} (p_i^{(\mu)} - p_i^{(\nu)})^2$$

and for difference between frequencies of heterozygote A_{ij}:

$$(1.11) \qquad \begin{aligned} \bar{A}_{ij} - \bar{A}_{ij}^{*} &= 2\sum_{\nu=1}^{g} \lambda^{(\nu)} \sum_{\nu=1}^{g} \lambda^{(\nu)} p_i^{(\nu)} p_j^{(\nu)} - 2\sum_{\nu=1}^{g} \lambda^{(\nu)} p_i^{(\nu)} \sum_{\nu=1}^{g} \lambda^{(\nu)} p_j^{(\nu)} \\ &= 2\sum_{\mu<\nu} \lambda^{(\mu)} \lambda^{(\nu)} (p_i^{(\mu)} - p_i^{(\nu)})(p_j^{(\mu)} - p_j^{(\nu)}) \qquad (i \neq j). \end{aligned}$$

The relations (1.10) concerning homozygotes contain a very striking content. In fact, we can immediately deduce

$$(1.12) \qquad \bar{A}_{ii}^{*} \leq \bar{A}_{ii} \qquad (i = 1, \ldots, m),$$

which are really based on the well-known *Cauchy's inequality.*
Moreover, it is evident that the equality sign in (1.12) can appear
if and only if $p_i^{(\nu)}(\nu = 1, \ldots, g)$ are all identical.

We thus have obtained the following result:

*In any population composed of some races in a mosaic manner
each of which possesses an equilibrium distribution, any homozygote
will have a frequency in the initial distribution not less than that in
the limit equilibrium distribution. Moreover, the former is actually
greater than the latter, unless all the component-races possess an
identical distribution.*

With regard to heterozygotes, since the right-hand members
of the relations (1.11) do not possess a definite sign, an analogous
result does not hold.

As an illustrative example we consider *ABO* human blood type.
In this case, the frequency of each homozygote *OO*, *AA*, *BB* will
diminish, compared with that in the initial distribution, when the
equilibrium state will be reached. In particular, homozygote *OO*
being composed of recessive genes and hence its frequency re-
presenting exactly that of phenotype *O*, the same is also valid for
the latter. The analogous result applies also to the phenotypes *M*
and *N* in *MN* blood type and the phenotype *q* in *Q* blood type.

2. Anthropological applications.

The relations (1.7), combined with (1.4), further implies

$$(2.1) \qquad \sum_{\nu=1}^{g} \lambda^{(\nu)}(p_i^{(\nu)} - p_i) = 0 \qquad (i = 1, \ldots, m).$$

These relations being homogeneous in $\lambda^{(\nu)}$, only their ratios
are essential, namely, the accessory condition (1.4) may now be
removed. These m relations may be regarded as a system of linear
equations to determine the ratios between $\lambda^{(\nu)}(\nu = 1, \ldots, g)$. To
state more precisely, m relations (2.1) can be considered as
simultaneous equations to determine the rate $\lambda^{(\nu)}(\nu = 1, \ldots, g)$ of
each sub-race $X^{(\nu)}$ possessing the distribution $\{p_i^{(\nu)}\}$ when the dis-
tribution $\{p_i\}$ of the composed population X is also known.

As these equations are all homogeneous in $\lambda^{(\nu)}$, there exist
essentially $g-1$ unknown quantities expressed as the ratios between
them. On the other-hand, because of the fundamental relations

$$(2.2) \qquad \sum_{i=1}^{m} p_i^{(\nu)} = \sum_{i=1}^{m} p_i = 1 \qquad (\nu = 1, \ldots, g),$$

the sum of all the left-hand sides of these m equations vanishes
identically, and hence the system (2.1) consists of at most $m-1$
independent equations.

We first consider the case $m > g$. A system of $m-1$ equations

in $g-1$ unknown quantities has then in general no solution. In order that the system is solvable, it is necessary and sufficient that the rank ρ, say, of (m, g) matrix

(2.3)
$$
\begin{pmatrix}
p_1^{(1)} - p_1 & \cdots & p_1^{(g)} - p_1 \\
\cdots & \cdots & \cdots \\
p_m^{(1)} - p_m & \cdots & p_m^{(g)} - p_m
\end{pmatrix}
$$

does not exceed $g-1$. But, in such a case, as we are now considering, where the relations in question hold without any contradiction, this condition must also be supposed valid. Furthermore, we shall have ordinarily $\rho = g-1$ and hence the ratios between $\lambda^{(\nu)}$ will be uniquely determined by the equations (2.1). Suppose, then, that for instance the first $g-1$ columns are independent, the required ratios are given by

$$
\lambda^{(1)} : \cdots : \lambda^{(\nu)} : \cdots : \lambda^{(g)} = -\Pi^{(1)} : \cdots : (-1)^{\nu}\Pi^{(\nu)} : \cdots : (-1)^{g}\Pi^{(g)},
$$

where, the $(g-1, g)$ sub-matrix composed of the first $g-1$ rows of (2.3) being denoted by

(2.4)
$$
\begin{pmatrix}
p_1^{(1)} - p_1 & \cdots & p_1^{(g)} - p_1 \\
\cdots & \cdots & \cdots \\
p_{g-1}^{(1)} - p_{g-1} & \cdots & p_{g-1}^{(g)} - p_{g-1}
\end{pmatrix}
$$

$\Pi^{(\nu)}$ expresses the determinant of the matrix obtained from this sub-matrix by deleting the νth column, i.e.

(2.5)
$$
\Pi^{(\nu)} = \begin{vmatrix}
p_1^{(1)} - p_1 & \cdots p_1^{(\nu-1)} - p_1 & p_1^{(\nu+1)} - p_1 & \cdots p_1^{(g)} - p_1 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
p_{g-1}^{(1)} - p_{g-1} & \cdots p_{g-1}^{(\nu-1)} - p_{g-1} & p_{g-1}^{(\nu+1)} - p_{g-1} & \cdots p_{g-1}^{(g)} - p_{g-1}
\end{vmatrix}
$$
$$
(\nu = 1, \ldots, g).
$$

The result remains valid also in case $m = g$.

We next consider the contary case $\rho < g-1$. The system (2.1) is then always solvable, but the solution is not determined uniquely. For instance, always so it is in case $m < g$. Such circumstances are evidently owing to the fact that the number of component-races exceeds that of genes and hence the rates of mixture giving the same distribution as in the composed population are possible in infinitely many manners. Even though $m = g$, we have still $\rho < g-1$, provided $p_i^{(\mu)} = p_i^{(\nu)}(i = 1, \ldots, m)$ for some pair of μ, ν. This is the case where the both races $X^{(\mu)}$ and $X^{(\nu)}$ possess an identical distribution, and hence any change in rate of mixture between them alone has no effect upon the total distribution; the indefiniteness of the rate of mixture is thus a quite evident consequence in this case.

The above stated result will now be concretely illustrated by ABO blood type. In this case, we have $m = 3$. If there exist

more than three sub-races, the rates of mixture are determined not uniquely and hence the case $g > m$ may be set aside as an uninteresting one.

We first consider the case $g = 3$, i.e. the case where three sub-races $X^{(1)}$, $X^{(2)}$ and $X^{(3)}$ are existent. Denoting by p, q, r, as usual, the frequencies of genes A, B, O respectively, and making use of analogous suffix-notations as in general consideration, the system of equations in question become

$$\lambda^{(1)}(p^{(1)}-p) + \lambda^{(2)}(p^{(2)}-p) + \lambda^{(3)}(p^{(3)}-p) = 0,$$
$$(2.6) \qquad \lambda^{(1)}(q^{(1)}-q) + \lambda^{(2)}(q^{(2)}-q) + \lambda^{(3)}(q^{(3)}-q) = 0,$$
$$\lambda^{(1)}(r^{(1)}-r) + \lambda^{(2)}(r^{(2)}-r) + \lambda^{(3)}(r^{(3)}-r) = 0.$$

Among these relations the last is not independent of the first two, from which we obtain

$$(2.7)\ \lambda^{(1)} : \lambda^{(2)} : \lambda^{( )} = \begin{vmatrix} p^{(2)}-p & p^{(3)}-p \\ q^{(2)}-q & q^{(3)}-q \end{vmatrix} : \begin{vmatrix} p^{(3)}-p & p^{(1)}-p \\ q^{(3)}-q & q^{(1)}-q \end{vmatrix} : \begin{vmatrix} p^{(1)}-p & p^{(2)}-p \\ q^{(1)}-q & q^{(2)}-q \end{vmatrix}.$$

If in particular all terms in the right-hand side of (2.7) vanish simultaneously, the ratios connot be determined uniquely. This is a special case, in which the relations

$$(p^{(1)}-p)/(q^{(1)}-q) = (p^{(2)}-p)/(q^{(2)}-q) = (p^{(3)}-p)/(q^{(3)}-q)$$

hold good and hence the rank of the coefficient-matrix in original equations is less than $g-1 = 2$. Except this special case only, the ratios $\lambda^{(1)} : \lambda^{(2)} : \lambda^{(3)}$ are uniquely determined such as they stand in (2.7).

We next turn to the case where there exist two sub-races $X^{(1)}$ and $X^{(2)}$. The corresponding system of equations then becomes

$$\lambda^{(1)}(p^{(1)}-p) + \lambda^{(2)}(p^{(2)}-p) = 0,$$
$$\lambda^{(1)}(q^{(1)}-q) + \lambda^{(2)}(q^{(2)}-q) = 0,$$
$$\lambda^{(1)}(r^{(1)}-r) + \lambda^{(2)}(r^{(2)}-r) = 0,$$

the last equation being here also dependent. Since this is the case with $m = 3 > 2 = g$, in order that the equations are consistent, the coefficient-determinant obtained from the first two must necesserily vanish, i.e. $(p^{(1)}-p)/(q^{(1)}-q) = (p^{(2)}-p)/(q^{(2)}-q)$. Then we must evidently have further relations

$$(p^{(1)}-p)/(p^{(2)}-p) = (q^{(1)}-q)/(q^{(2)}-q) = (r^{(1)}-r)/(r^{(2)}-r).$$

Conversely, if the last condition is satisfied, the rate of mixture is given by

$$\frac{\lambda^{(1)}}{\lambda^{(2)}} = \frac{p^{(2)}-p}{p-p^{(1)}} \left(= \frac{q^{(2)}-q}{q-q^{(1)}} = \frac{r^{(2)}-r}{r-r^{(1)}} \right).$$

There exists, however, only an exceptional case where

$$p^{(1)} = p^{(2)} = p, \quad q^{(1)} = q^{(2)} = q, \quad r^{(1)} = r^{(2)} = r.$$

In this case the numerator and the denominator in the right-hand of the above relation vanish. But, as noticed generally, this is the case where the composed population is born of two races with the same distribution. That the rate of mixture cannot be determined from the given data only, is a matter of course.

We now point out an anthrological significance of the relations obtained above for case $g = 3$. Suppose that three races $X^{(1)}$, $X^{(2)}$, $X^{(3)}$ are mixed in certain rates long ago and constituted a population X, but one of the sub-races, $X^{(3)}$ say, has ruined in mutable course of long history, and the remaining two sub-races $X^{(1)}$ and $X^{(2)}$ are actually surviving sepatately from the composed population. Then, if the distributions of genes in $X^{(1)}$, $X^{(2)}$, $X^{(3)}$ and X are all known, the above relations may be considered to give the rates of mixture. However, the same relations imply a further interesting application. In fact, suppose now that the distribution of the ruined sub-race $X^{(3)}$ is unkown although the rates of mixture between three sub-races is known; such a case will happen rather often as a paractical problem. We can then estimate the distribution of the ruined sub-race inversely. For this purpose we have only to solve the equations (2.7) or the equivalent equations (2.6) with respect to $p^{(3)}$, $q^{(3)}$ and $r^{(3)}$. Suppose again $\lambda^{(1)} + \lambda^{(2)} + \lambda^{(3)} = 1$, for the sake of brevity, we thus have

$$\text{(2.8)} \quad \begin{aligned} p^{(3)} &= (p - \lambda^{(1)}p^{(1)} - \lambda^{(2)}p^{(2)})/\lambda^{(3)}\,, \\ q^{(3)} &= (q - \lambda^{(1)}q^{(1)} - \lambda^{(2)}q^{(2)})/\lambda^{(3)}\,, \\ r^{(3)} &= (r - \lambda^{(1)}r^{(1)} - \lambda^{(2)}r^{(2)})/\lambda^{(3)}\,. \end{aligned}$$