86. Probability-theoretic Investigations on Inheritance.

II_2. 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.)

3. Inhomogeneous distribution.

A population with a mosaic composition, discussed in two preceding sections, may be regarded as a sort of population with *inhomogeneous*, i.e. *non-uniform* distribution. Let us now consider a general type of inhomogeneous distributions. Now, generally in inheritance phenomena, the unit of observations is an individual. Hence, if we attempt to introduce the concept of distribution-density of various quantities, it will be reprehensible to divide the region immoderately to pieces. But, as in case of usual population-density, we may define the concept of *distribution-density* suitably according to such a moderate division that the probability-theoretic considerations are expected to apply appropriately.

We consider a region G in which the inheritance phenomenon with genes $A_i (i = 1, \dots, m)$ will be observed. Suppose that G is a region on a plane and the position of each point of G is denoted by coordinates (x, y) with respect to a suitably chosen rectanular coordinate system.——In case where G is a region on a surface, the subsequent discussion will also remain valid with slight modifications; in fact, we have only to choose a suitable curvilinear coordinate system and correspondingly to make use of its surface element instead of $dxdy$.

Let the distribution-density of the phenotypes A_{ij} in the region G be denoted (at least approximately) by a set of contiuous or, more generally, piecewise continuous functions

$$P_{ij}(x, y) \qquad (i, j = 1, \dots, m; i \leqq j) \tag{3.1}$$
$$(P_{ij}(x, y) = P_{ji}(x, y)).$$

Let the distribution-density of the genes A_i be denoted by

$$p_i(x, y) \qquad (i = 1, \dots, m). \tag{3.2}$$

Then, at each point of G, the fundamental relations

$$\sum_{i=1}^{m} p_i(x, y) = 1, \qquad \sum_{i \leqq j} P_{ij}(x, y) = 1 \tag{3.3}$$

must hold. In any small part the distribution being regarded to be in an equilibrium state, we have the interrelations

$$P_{ii}(x, y) = p_i(x, y)^2, \quad P_{ij}(x, y) = 2p_i(x, y)p_j(x, y) \quad (i \neq j), \tag{3.4}$$

whence we deduce

$$(3.5)\qquad p_i(x,y)=P_{ii}(x,y)+\tfrac{1}{2}\sum_{j\neq i}P_{ij}(x,y)\quad (i=1,\dots,m)\,,$$

corresponding to formulae (2.9) in I.

Let further the population-density in G be denoted by $\delta(x,y)$. The distribution of genotypes in any sub-region G' is then given by

$$(3.6)\qquad P_{ij}[G']=\iint_{G'}P_{ij}(x,y)\delta(x,y)dx\,dy\Big/\iint_{G'}\delta(x,y)dxdy\qquad (i\leqq j)\,;$$

it is of course permissible that G' is not a domain but may consist of a certain number of connected pieces like colonies. In particular, the distribution of phenotypes in the whole reigion G is expressed by

$$(3.7)\qquad P_{ij}\equiv P_{ij}[G]=\frac{1}{\varDelta}\iint_G P_{ij}(x,y)\delta(x,y)dxdy\quad (i\leqq j)\,,$$

$\varDelta$ denoting the total population of G, i.e.

$$(3.8)\qquad \varDelta=\iint_G\delta(x,y)dxdy\,.$$

The relation (3.7) is a generalization of (1.5) of II; here the $\delta(x,y)dxdy/\varDelta$ appears corresponding to $\lambda^{(\nu)}$ in (1.5) of II.

The condition that the distribution is homogeneous in a subregion G' expresses the fact that the density function $p_i(x,y)$ remains there identically constant. Hence, a population consisting, in mosaic manner, of g subraces each of which possesses an equilibrium distribution may be regarded essentially as a special case where every $p_i(x,y)(i\leqq j)$ becomes a step-function taking at most g different values.

We now consider the limit distribution which will be reached by the process of buffer effect (cf. §4 of I). We have then, in quite similar manner as in §1 of II, the limit equilibrium distribution given by

$$(3.9)\qquad \begin{aligned} p_i &= P_{ii}+\tfrac{1}{2}\sum_{j\neq i}P_{ij}\\ &=\frac{1}{\varDelta}\iint_G\Big(P_{ii}(x,y)+\tfrac{1}{2}\sum_{j\neq i}P_{ij}(x,y)\Big)\delta(x,y)dxdy\\ &=\frac{1}{\varDelta}\iint_G p_i(x,y)\delta(x,y)dxdy\quad (i=1,\dots,m)\end{aligned}$$

whence the frequencies of genotypes in the limit distribution

$$(3.10)\qquad \begin{aligned} P^*_{ii}&\equiv p_i^2=\frac{1}{\varDelta^2}\Big(\iint_G p_i(x,y)\delta(x,y)dxdy\Big)^2,\\ P^*_{ij}&\equiv 2p_ip_j=\frac{2}{\varDelta^2}\iint_G p_i(x,y)\delta(x,y)dxdy\iint_G p_j(x,y)\delta(x,y)dxdy\end{aligned}\qquad (i\neq j).$$

On the other hand, in view of (3.7) and (3.4), the original distribution of genotypes will be given by

$$
\begin{aligned}
&P_{ii} = \frac{1}{\Delta}\iint_G p_i(x, y)^2\delta(x, y)dxdy, \\
(3.11)\qquad & \\
&P_{ij} = \frac{2}{\Delta}\iint_G p_i(x, y)p_j(x, y)\delta(x, y)dxdy \quad (i \neq j).
\end{aligned}
$$

Comparing the original distribution with its limit distribution, we deduce, by taking into account of (3.8), the relations

$$(3.12)\qquad P^*_{ii} \leq P_{ii} \qquad (i = 1, \dots, m),$$

which are based on the well-known *Schwarz's integral inequality*. More precisely, we obtain the relations corresponding to (1.10) and (1.11) of II, namely

$$
\begin{aligned}
P_{ii}-P^*_{ii} &= \frac{1}{\Delta^2}\Big\{\iint_G \delta(x, y)dxdy\iint_G p_i(x, y)^2\delta(x, y)dxdy \\
&\qquad -\Big(\iint_G p_i(x, y)\delta(x,y)dxdy\Big)^2\Big\} \\
&= \frac{1}{2\Delta^2}\Big\{\iint_G \delta(\xi, \eta)\, d\xi\, d\eta\iint_G p_i(x, y)^2\delta(x, y)dxdy \\
(3.13)\qquad &\qquad +\iint_G \delta(x, y)dxdy\iint_G p_i(\xi, \eta)^2\delta(\xi, \eta)d\xi d\eta \\
&\qquad -2\iint_G p_i(x, y)\,\delta\,(x, y)dxdy\iint_G p_i(\xi, \eta)\delta(\xi, \eta)d\xi d\eta\Big\} \\
&= \frac{1}{2\Delta^2}\iint_G\iint_G (p_i(x, y)-p_i(\xi, \eta))^2\delta(x, y)\delta(\xi, \eta)dxdy\, d\xi d\eta
\end{aligned}
$$

and

$$
\begin{aligned}
P_{ij}-P^*_{ij} &= \frac{2}{\Delta^2}\Big\{\iint_G \delta(x, y)dxdy\iint_G p_i(x, y)p_j(x, y)\delta(x, y)dxdy \\
&\qquad -\iint_G p_i(x, y)\delta(x, y)dx\,dy\iint_G p_j(x, y)\delta(x, y)dxdy\Big\} \\
(3.14)\qquad &= \frac{1}{\Delta^2}\iint_G\iint_G (p_i(x, y)-p_i(\xi, \eta))(p_j(x, y)-p_j(\xi, \eta)) \\
&\qquad\qquad \cdot\delta(x, y)\delta(\xi, \eta)dxdy\, d\xi d\eta.
\end{aligned}
$$

The above inequality (3.12) is an immediate consequence of the relation (3.13). Moreover, it is also obviously seen that the equality sign in (3.12) can occur if and only if the functions $p_i(x, y)$ reduce to constants, i.e.

$$p_i(x, y) \equiv p_i \qquad (i = 1, \dots, m;\ (x, y) \in G).$$

We thus obtained the following result:

In any population with distribution variable with respect to positions, any homozygote will have a frequency in initial distribution not less than that in limit equilibrium distribution. Moreover, the former is actually greater than the latter, unless the initial distribution is uniform everywhere.

An equivalent statement may also be formulated as follows:

A frequency of any homozygote observed in a region with an inhomogeneous distribution will always be estimated too large in comparison with frequency expected from corresponding frequency of genes in the limit distribution.

Any phenotype consisting of a single kind of recessive gene will also be subject to the same statement. The illustrative examples may be offered by the phenotypes *O; M, N; q* of *ABO; MN; Q* human blood types respectively.