29. *Probability-theoretic Investigations on Inheritance. VII*$_5$. *Non-Paternity Problems.*

By Yûsaku KOMATU.

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

(Comm. by T. FURUHATA, M.J.A., Feb. 12, 1952.)

6. Discontinuity on probablity based upon recessive genes.

As remarked in § 1, the probability of proving non-paternity is, in general, less in case upon phenotypes alone than in case upon the corresponding genotypes, provided that there exist recessive genes. Besides this trivial fact, we notice here a further remarkable fact which has probably not yet been explicitly noticed. We shall discuss in the present section this apparently curious phenomenon, which may be called the *discontinuity on probability of proving non-paternity based upon recessive genes.*

Now, if, in the formula (5.3), representing the whole probability of proving non-paternity with the aid of *ABO* blood type, we put $r=0$ and correspondingly $p+q=1$, then it becomes

$$\varphi \equiv [P_{ABO}]^{r=0} = qp^4 + pq^4 = pq(p^3+q^3) = pq(1-3pq). \tag{6.1}$$

On the other hand, if, in the corresponding formula (5.1), representing that with the aid of *MN* blood type, we merely substitute p and q instead of s and t respectively, then it becomes

$$\psi \equiv [P_{MN}]^{(s,t)=(p,q)} = pq(1-pq). \tag{6.2}$$

Hence, φ is less than ψ in general, i.e., surely so unless pq vanishes; in fact, the difference is equal to

$$\varphi - \psi = -2p^2q^2 \leqq 0. \tag{6.3}$$

In general, the comparison between the modes of inheritance of *ABO* and *MN* blood types immediately shows that, if particularly the gene *O* would be lacking, then the mode of inheritance of the former consisting of both genes *A* and *B* alone would reduce essentially to that of the latter. Consequently, it would superficially be expected as plausible that, if, in any general result on *ABO* blood type, we put particularly $r=0$ and replace s, t instead of p, q respectively, then a corresponding result on *MN* blood type would be obtained. Although this is really the case for the most part, all does not go well. A counter-example is offered by non-paternity problem under consideration. If r vanishes, the probability P_{ABO}

increases discontinuously and attains the greater value ψ given in (6.2) than the limiting value φ of P_{ABO} for $r \to 0$ given in (6.1); P_{ABO} being a polynomial, it is, of course, a continuous function of its arguments. Thus, the whole probability of proving non-paternity with the aid of *ABO* blood type must, in the extreme case $r=0$, i.e., in case of complete lack of the gene *O*, be expressed by (6.2), not by (6.1).

The reason why such a discontinuity occurs, may be illustrated as follows. If the gene *O* would be lacking, then the phenotypes *A* and *B* would consist of respective homozygote *AA* and *BB* alone. Hence, for instance, given a mother-child combination (*A*; *A*), a man *B* would be proved as being not a true father. But, if the gene *O* is existent at any rate, however small, the phenotypes *A* and *B* contain then generally, besides homozygotes, also heterozygotes *AO* and *BO*, respectively. Consequently, given a mother-child combination (*A*; *A*), although a man *BB* cannot be a true father, a man *BO* may possibly be a true father; in fact, the mating between *BO* and *AA* or *AO* can produce a child *AO*. Since the genotypes *BB* and *BO* cannot be distinguished upon phenotypes alone, it is impossible, against the given mother-child combination, to establish the non-paternity of the whole of the type *B* upon phenotypes alone. Thus and similarly, putative fathers *B*, *A*; *B* and *A* against mother-child combinations (*A*; *A*), (*B*; *B*); (*AB*; *A*) and (*AB*; *B*) respectively, which would be deniable in case of complete lack of the gene *O*, must be dismissed undecidedly if the gene *O* is existent. The sum of sub-probabilities corresponding to these four pairs in case of lacking *O* is equal to

$$p^3q^2+q^3p^2+p^2qq^2+pq^2p^2=2p^2q^2(p+q)=2p^2q^2. \tag{6.4}$$

It is a matter of course that the deficiency (6.3) cancels exactly with the sum given in (6.4).

Because of respective modes of inheritance, such a discontinuity does *not* occur between A_1A_2BO and *ABO* blood types. In fact, if in $P_{A_1A_2BO}$ we put $(p_1, p_2)=(p, 0)$ or $(p_1, p_2)=(0, p)$, then it reduces precisely to P_{ABO}; cf. (5.7). Similarly, $P_{Qq\pm}$ reduces to P_Q, by putting $(v_1, v_2)=(v, 0)$ or $(v_1, v_2)=(0, v)$; cf. (5.10) and (5.5).

By the way, we notice here that, if the gene *B* is lacking, the mode of inheritance of A_1A_2BO blood type becomes essentially the same as that of $Qq_\pm$ blood type; similarly, if one and only one of the genes *A* and *B* is lacking, mode of inheritance of *ABO* blood type becomes essentially the same as that of *Q* blood type. But, for such cases, the discontinuity of the above mentioned nature does *never* occur, as really shows the respective comparison.

With respect to the probabilities on mixed combinations, the circumstances are also quite similar.

7. Distribution of maximum probability.

In preceding sections we have derived the probabilities of proving non-paternity in various cases. We now, consider the problem of determining such a distribution in each case that the respective probability attains its maximum value.

We first observe, as a model, the simplest case realized by *MN blood type*. Its probability is given by (5.1), i.e.,

$$P_{MN} = st(1-st). \tag{7.1}$$

Since both arguments s and t are dependent and, in fact, they are connected by the relation $s+t=1$, the quantity P_{MN} is essentially a function of either of them, say s alone, ranging over the interval $0 \leqq s \leqq 1$. But, we may rather regard it as a function of the product $x=st$. In view of the relation $x=((s+t)^2-(s-t)^2)/4=(1-(s-t)^2)/4$, the variable x ranges over the interval $0 \leqq x \leqq 1/4$. Thus, differentiating $P_{MN}=x(1-x)$ with respect to x, it becomes $dP_{MN}/dx=1-2x$ and hence P_{MN} increases strictly with x surely for $0 \leqq x \leqq 1/4$. Therefore, P_{MN} attains its maximum if and only if $x=1/4$. It is evident that x becomes equal to $1/4$ if and only if $s=t$. We thus conclude that the maximum of P_{MN} is attained if and only if the distribution of genes is given by

$$s = t = 1/2, \tag{7.2}$$

and the maximum value is equal to

$$(P_{MN})^{\max} = 3/16 = 0.1875. \tag{7.3}$$

The distribution of phenotypes corresponding to (7.2) is, of course,

$$\overline{M}=\overline{N}=1/4=0.25, \qquad \overline{MN}=1/2=0.5. \tag{7.4}$$

This is a well-known result[1].

We next consider the case of mixed combinations given by (5.2), i.e.,

$$P'_{MN} = s't'(1-st). \tag{7.5}$$

A few different circumstances appear. In fact, P'_{MN} may be regarded as a function of two independent variables $x'=s't'$ and $x=st$, and moreover it possesses a form being separate with respect to them, i.e., it is a product of two factors each of which is a function of one variable alone. Hence, in order to maximize P'_{MN}, it suffices only to maximize each factor separately. Thus, we get the maxi-

1) Cf., for instance, loc. cit.[1] of VII$_4$.

mizing distribution of genes and the corresponding maximum value:

$$s' = t' = 1/2\,; \qquad st = 0\,; \tag{7.6}$$

$$(P'_{MN})^{\max} = 1/4 = 0.25\,. \tag{7.7}$$

The distribution of phenotypes corresponding to (7.6) is then

$$\begin{aligned} &\overline{M}'=\overline{N}'=1/4=0.25, \qquad \overline{MN}'=1/2=0.5; \\ &\overline{M}=\overline{MN}=0,\ \overline{N}=1 \quad \text{or} \quad \overline{N}=\overline{MN}=0,\ \overline{M}=1. \end{aligned} \tag{7.8}$$

By comparing the result given in (7.3) with that in (7.7), we see that the probability with respect to pure combination is 18.75% even in the most favorable case, while that with respect to mixed combination can amount to 25%. But, the extremal distribution (7.8) for the latter being extremely special, it will not be practically realized even approximately. On the other hand, it will be noticed, as shows a general view, that most of the practically realized distributions are near to that given in (7.4); hence the realized probabilities will also be near to the maximum value given by (7.3).

We now proceed to consider the general case given by (2.20), i e.,

$$P=1-2S_2+S_3-2S_2^2+2S_4+3S_2S_3-3S_5. \tag{7.9}$$

The problem is to maximize this quantity under accessory conditions

$$0 \leqq p_i \ (i=1, \ldots, m), \qquad \sum_{i=1}^{m} p_i=1. \tag{7.10}$$

By means of a general method due to Lagrange on extemum problem with accessory conditions, we obtain a system of equations for determining the values of $p_i\,(i=1, \ldots, m)$ which maximize P:

$$\frac{\partial}{\partial p_i}\left(P-\lambda\left(\sum_{j=1}^{m} p_j-1\right)\right)=0 \ (i=1, \ldots, m), \quad \sum_{i=1}^{m} p_i=1; \tag{7.11}$$

λ denoting the so-called *Lagrangean multiplier*. The first m equations become

$$p_i(-2+3p_i-8S_2+8p_i^2+6S_3+9p_iS_2-15p_i^3)=\lambda \quad (i=1, \ldots, m). \tag{7.12}$$

From a symmetric character of the system of equations, it is evidently satisfied by the set of values

$$p_i = 1/m \qquad (i=1, \ldots, m). \tag{7.13}$$

Since $S_\nu=1/m^{\nu-1}$ for (7.13), the corresponding stationary value of P is equal to

$$(P)^{\text{stat}} = (1-1/m)(1-1/m-2/m^2+3/m^3). \tag{7.14}$$

It would, moreover, be believed that the value in (7.14) would yield perhaps the actual maximum, as really the case, in particular, for $m=2$. The distribution of genotypes corresponding to (7.13) is then

$$(7.15) \qquad \bar{A}_{ii}=1/m^2, \qquad \bar{A}_{ij}=2/m^2 \qquad (i,j=1,\ldots,m;\ i<j).$$

Now, differentiating the stationary value given in (7.14) with respect to $1/m$ regarded as if a continuous parameter, then we get

$$\frac{d}{d(1/m)}(P)^{\text{stat}}$$
$$=-\left(1-\frac{1}{m}\right)\left(1-\frac{1}{m^2}+\frac{4}{m}\left(1-\frac{2}{m}\right)\right)-\left(\left(1+\frac{1}{m}\right)\left(1-\frac{2}{m}\right)+\frac{3}{m^2}\right).$$

It is evident that the right-hand side of the last expression remains steadily negative provided $1/m \leqq 1/2$. Hence, the stationary value (7.14) increases monotone with $m\,(\geqq 2)$, and tends asymptotically to the limit 1 as $m \to \infty$:

$$(7.16) \qquad \lim_{m\to\infty}(P)^{\text{stat}} = 1.$$

We thus conclude that probability of proving non-paternity in the most favorable distribution becomes nearer to $1=100\%$, the more the number of allelomorphic genes is, and, moreover, tends to 1 as $m \to \infty$. This seems to be a remarkable fact. For instance, the values of $(P)^{\text{stat}}$ in (7.14) are 0.1875, 0.3704, 0.5039, 0.5952, 0.7947 and 0.9799 for $m=2$, 3, 4, 5, 10 and 100, respectively.

—To be continued—