30. *Probability-theoretic Investigations on Inheritance. VII$_6$. 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.)

7^{bis}. Distribution of maximum probability.

We consider the case of mixed combination given by (4.12), i.e.,

$$P'=1-2S'_2+S'_3-2S_{1,1}^2+2S_{2,2}+3S_{1,1}S_{1,2}-3S_{2,3}. \tag{7.17}$$

The problem is to maximize this quantity under accessory conditions

$$0\leqq p_i,\ p'_i \quad (i=1,\ldots,m); \qquad \textstyle\sum_{i=1}^m p_i=\sum_{i=1}^m p'_i=1. \tag{7.18}$$

The set of maximizing distributions $\{p_i\}$ and $\{p'_i\}$, if existent interior to the ranges, would be determined by a system of equations

$$\left.\begin{aligned}&\frac{\partial}{\partial p_i}\Big(P'-\lambda\Big(\sum_{j=1}^m p_j-1\Big)-\lambda'\Big(\sum_{j=1}^m p'_j-1\Big)\Big)=0,\\ &\frac{\partial}{\partial p'_i}\Big(P'-\lambda\Big(\sum_{j=1}^m p_j-1\Big)-\lambda'\Big(\sum_{j=1}^m p'_j-1\Big)\Big)=0\end{aligned}\right\}\quad (i=1,\ldots,m),$$

$$\sum_{i=1}^m p_i=\sum_{i=1}^m p'_i=1;$$

λ and λ' denoting the Lagrangean multipliers. The first $2m$ equations become

$$p'_i(-4S_{1,1}+4p_ip'_i+3S_{1,2}+3p'_iS_{1,1}-6p_ip'^2_i)=\lambda,$$
$$-2p'_i+3p'^2_i-4p_iS_{1,1}+4p_i^2p'_i+3p_iS_{1,2}+6p_ip'_iS_{1,1}-9p_i^2p'^2_i=\lambda'$$
$$(i=1,\ldots,m).$$

However, as suggested by the previously discussed special case $m=2$, it seems that the maximum of P' will rather be attained by an extreme distribution of $\{p_i\}$ lying on the boundary of its range; namely,

$$p_i=1 \quad (i=i_0), \qquad p_i=0 \quad (i\neq i_0) \tag{7.19}$$

for any value of $i_0\,(1\leqq i_0\leqq m)$. For such a distribution, P' becomes

$$P'_*=1-2S'_2+S'_3, \tag{7.20}$$

the value being independent of i_0.

The maximum of P'_* under the condition $\sum p'_i=1$ is surely attained by the symmetric distribution

$$p'_i=1/m \qquad (i=1,\ldots,m). \tag{7.21}$$

In fact, by means of the usual method, the set $\{p'_i\}$ maximizing P'_* is determined by a system of equations

$$(7.22)\qquad \begin{aligned} 0 &= (\partial/\partial p'_i)(P'_* - \lambda'_* (\textstyle\sum_{j=1}^m p'_j - 1)) \\ &= -4p'_i + 3p'^2_i - \lambda'_* \quad (i=1,\dots,m), \end{aligned} \qquad \textstyle\sum_{i=1}^m p'_i = 1$$

λ'_* being a multiplier. The difference of the ith and the jth equations becomes

$$(p'_i - p'_j)(4 - 3(p'_i + p'_j)) = 0.$$

Because of the restriction $p'_i + p'_j \leqq 1$ $(i \neq j)$, we conclude that the relation $p'_i = p'_j$ must hold for every pair of i and j. Hence, the maximizing distribution for P'_* is indeed given by (7.21); the value of the multiplier is then equal to $\lambda'_* = -(4m-3)/m^2$. Thus, we get

$$(7.23)\qquad (P'_*)^{\max} = 1 - 2/m + 1/m^2 = (1 - 1/m)^2.$$

It is evident that the right-hand side of the last expression increases with m and tends asymptotically to unity as $m \to \infty$. Its values are 0.25, 0.4444, 0.5625, 0.64, 0.81 and 0.9801 for $m=2$, 3, 4, 5, 10 and 100, respectively.

It seems most likely that the maximum of P'_* just obtained is simultaneously that of P' in (7.17). At any rate it is sure that the inequality holds good:

$$(7.24)\qquad (P')^{\max} \geqq (P'_*)^{\max} = (1 - 1/m)^2.$$

Comparing the both relations (7.24) and (7.14), we get

$$(P')^{\max} - (P)^{\text{stat}} \geqq (1/m^2)(1 - 1/m)(2 - 3/m),$$

the right-hand side of which is steadily positive provided $m \geqq 2$. This is quite a reasonable fact. In fact, P' reduces to P when the distribution $\{p'_i\}$ coincides particularly with $\{p_i\}$. Hence, the degrees of freedom with respect to the variables are greater in case of P' than in case of P, what implies immediately the inequality $(P')^{\max} \geqq (P)^{\max}$.

In case of *ABO blood type*, the result on maximizing distribution is classical[1]. In fact, the probability given by (5.3) has to be regarded as a function of two independent variables, e.g., p and q, based upon the identity $r = 1 - p - q$. Differentiation of thus obtained function

$$P_{ABO} = p(1-p)^4 + q(1-q)^4 + pq(1-p-q)^2(2+p+q)$$

with respect to p and to q leads to the pair of equations

1) Cf., for instance, loc. cit.[1] of VII_4; or also loc. cit.[2] of VII_4.

$$\begin{aligned}(7.25)\qquad 0=\partial P_{ABO}/\partial p&=(1-p)^3(1-5p)\\&\quad+q(1-p-q)(2-4p-q-4p^2-5pq-q^2),\\0=\partial P_{ABO}/\partial q&=(1-q)^3(1-5q)\\&\quad+p(1-p-q)(2-p-4q-p^2-5pq-4q^2),\end{aligned}$$

which yields, together with $r=1-p-q$, the maximizing distribution

$$(7.26)\qquad p=q=0.2212,\quad r=0.5576.$$

The extremal values of p and q coinciding each other are both the root of the quartic equation

$$(7.27)\qquad 25x^4-16x^3+9x^2-6x+1=0.$$

That this equation possesses a unique root contained in the interval $0<x<1/2$ can easily be verified, for instance, by means of the so-called *Strum's chain* in the theory of algebraic equations; it possesses a root also in the interval $1/2<x<1$ which does not satisfy the requirement of maximization.

The maximum of P_{ABO} corresponding to the distribution (7.26) becomes

$$(7.28)\qquad (P_{ABO})^{\max}=0.1999;$$

the maximizing distribution of phenotypes being

$$(7.29)\qquad \overline{O}=0.3109,\quad \overline{A}=\overline{B}=0.2956,\quad \overline{AB}=0.0979.$$

Now, the stationary value of P, given in (7.14), becomes in case $m=3$, as already stated, equal to $10/27=0.3704$. The value $(P_{ABO})^{\max}$ in (7.28) is nearly the half of this value. The quantity P expressing the probability in question with the aid of genotypes, this deficiency is no other than caused by the existence of a recessive gene, i.e., O.

We next consider the probability P'_{ABO} given in (5.4), concerning the mixed combination. This quantity reducing, for $(p', q', r')=(p, q, r)$, just to P_{ABO}, its maximum value is never less than the value given in (7.28). Moreover, since, for a particular pair of distributions

$$(7.30)\qquad p=q=0,\quad r=1;\quad p'=q'=r'=1/3,$$

P'_{ABO} becomes equal to 10/27, it is sure that the relation

$$(7.31)\qquad (P'_{ABO})^{\max}\geqq 10/27=0.3704$$

holds good. The value standing in the right-hand side of the last inequality coincides accidentally with that of $(P)^{\mathrm{stat}}$ in (7.14), for $m=3$, and is nearly the twice of the maximum value of P_{ABO}.

In case of *A_1A_2BO blood type*, the problem of determining the maximizing distribution will be somewhat troublesome. We omit here the detailed analysis. As noticed above, the value $(P)^{\text{stat}}$ in (7.14) becomes, in case $m=4$, equal to $129/256=0.5039$. Since, in case of A_1A_2BO blood type, dominance relations are really existent, the maximum value of $P_{A_1A_2BO}$ will be considerably less than the last mentioned value. But, on the other hand, since this blood type is a sub-division of ABO blood type, the maximum value of $P_{A_1A_2BO}$ is not less than the value in (7.28). We thus get a rough estimation

$$(7.32) \qquad 0.1919 \leqq (P_{A_1A_2BO})^{\max} \leqq 0.5039.$$

We next consider the case of *Q blood type*. The probability given in (5.5) may be written in the form

$$(7.33) \qquad P_Q = uv^4 = (1-v)v^4.$$

The maximizing distribution is determined by means of the equation $0=dP_Q/dv=v^3(4-5v)$, whence it follows that P_Q is maximized at the distribution

$$(7.34) \qquad u=1/5=0.2,\ v=4/5=0.8;\ \bar{Q}=9/25=0.36,\ \bar{q}=16/25=0.64;$$

the maximum value of the probability being

$$(7.35) \qquad (P_Q)^{\max}=(1/5)(4/5)^4=256/3125=0.0819.$$

In mixed case, the probability is given by (5.6), i.e.,

$$(7.36) \qquad P'_Q = v^2u'v'^2 = v^2(1-v')v'^2.$$

In order that P'_Q attains its maximum, it is necessary that v is equal to 1. The quantity P'_Q then becomes equal to $(1-v')v'^2$. Hence, we get the maximizing distributions and the maximum value:

$$(7.37) \qquad u=0,\quad v=1;\quad u'=1/3=0.3333,\quad v'=2/3=0.6667;$$

$$(7.38) \qquad \bar{Q}=0,\quad \bar{q}=1;\quad \bar{Q}'=5/9=0.5556,\quad \bar{q}'=4/9=0.4444;$$

$$(7.39) \qquad (P'_Q)^{\max}=(1/3)(2/3)^2=4/27=0.1481.$$

In conclusion, we consider the case of *$Qq_\pm$ blood type*. The probabilities given in (5.10) and (5.11) can be written in the respective forms

$$(7.40) \qquad P_{Qq\pm}=uv^4+v_1v_2^4=(1-v_1-v_2)(v_1+v_2)^4+v_1v_2^4,$$

$$(7.41) \qquad P'_{Qq\pm}=v^2u'v'^2+v_2^2v_1'v_2'^2=v^2(1-v_1'-v_2')(v_1'+v_2')^2+v_2^2v_1'v_2'^2.$$

Differentiation of (7.40), considered as a function of two independent variables v_1 and v_2, with respect to each of them leads to

$$(7.42) \qquad (v_1+v_2)^3(4-5(v_1+v_2))+v_2^4=(v_1+v_2)^3(4-5(v_1+v_2))+4v_1v_2^3=0,$$

the system of equations for determining the maximizing distribution. It can easily be solved. In fact, we get, by subtraction, $v_2=4v_1$ whence it follows

(7.43) $$125v_1^3(4-25v_1)+256v_1^4=0\,.$$

Thus, we get the maximizing distribution for $P_{Qq\pm}$ and the maximum value:

(7.44) $$v_1=500/2869,\ v_2=2000/2869;\ v=2500/2869,\ u=369/2869;$$

(7.45) $$\bar{Q}=0.2407,\qquad \bar{q}_-=0.2737,\qquad \bar{q}_+=0.4856;$$

(7.46) $$\begin{aligned}(P_{Qq\pm})^{\max}&=(369/2869)(2500/2869)^4+(500/2869)(2000/2869)^4\\&=22414062500000000/194382520709325349=0.1153.\end{aligned}$$

Lastly, the maximum of (7.41) is attained evidently when the extreme distribution $v_2=v=1$ does appear. The probability then becomes equal to $(1-v_1'-v_2')(v_1'+v_2')^2+v_1'v_2'^2$, which leads, by differentiation with respect to each of v_1' and v_2', to the system of equations

(7.47) $$(v_1'+v_2')(2-3(v_1'+v_2'))+v_2'^2=(v_1'+v_2')(2-3(v_1'+v_2'))+2v_1'v_2'=0.$$

Thus, we get the maximizing distributions for $P'_{Qq\pm}$ and the maximum value:

(7.48) $$v_2=v=1,\ v_1=u=0;\ v_1'=6/23,\ v_2'=12/23;\ v'=18/23,\ u'=5/23;$$

(7.49) $$\begin{aligned}&\bar{Q}=\bar{q}_-=0,\quad \bar{q}_+=1;\qquad \bar{Q}'=205/529=0.3875,\\&\bar{q}'_-=180/529=0.3403,\quad \bar{q}'_+=144/529=0.2722;\end{aligned}$$

(7.50) $$\begin{aligned}(P'_{Qq\pm})^{\max}&=(5/23)(18/23)^2+(6/23)(12/23)^2\\&=2484/12167=0.2042.\end{aligned}$$

By comparing the results (7.35), (7.39), (7.46) and (7.50), we notice that the relations hold:

(7.51) $$(P'_{Qq\pm})^{\max}>(P'_Q)^{\max}>(P_{Qq\pm})^{\max}>(P_Q)^{\max},$$

the equality sign being excluded everywhere, among which the weaker inequalities $(P'_{Qq\pm})^{\max}\geqq(P'_Q)^{\max}\geqq(P_Q)^{\max}$ and $(P'_{Qq\pm})^{\max}\geqq(P_{Qq\pm})^{\max}\geqq(P_Q)^{\max}$ are trivial and could be preassigned without any calculation.