Maximum-Likelihood Estimation of the Proportion of Nonpaternity

RICHARD F. POTTHOFF AND MAURICE WHITTINGHILL

Department of Statistics and Department of Zoology,
University of North Carolina, Chapel Hill.

THE USE OF GENETIC DATA to estimate the proportion of nonpaternity or extramarital illegitimacy has been discussed in recent writings (Li, 1961, p. 35; MacCluer and Schull, 1963). We will use the term "proportion of nonpaternity" to mean the proportion of children (λ) for whom the putative or supposed father is not the true (biological) father; this is essentially the same concept that is used by Li (who speaks of "illegitimacy" rather than "nonpaternity") and by MacCluer and Schull. As MacCluer and Schull point out, the proportion λ is of some interest to geneticists, since the results of any segregation analysis or linkage detection study would be vitiated if λ were very much larger than zero in the population under investigation. It would also appear that sociologists might be interested in a measurement technique which estimates the sociological variable λ solely from genetic data; because of its objectivity, such a technique would provide a valuable comparison with other possible techniques such as those involving questionnaires or interviews.

We suppose that a number (n) of combinations or trios—each consisting of a putative father, a mother, and a child—have been examined with respect to some genetic trait and that the phenotype of each member of each trio has been ascertained. The basic problem then is to estimate λ from such data. Provided that computations are not prohibitively difficult, the method of maximum likelihood would probably be favored as the best method of estimation because of its many desirable properties (see, e.g., Fraser, 1958, pp. 224–228; Cramér, 1946, pp. 498ff.). The purpose of this paper is to examine how the maximum-likelihood method can be used in different situations to estimate the proportion λ of nonpaternity.

In Li's brief example, he exhibits a simple and quick method (not maximum likelihood) for estimating λ in the case of a trait for which there are two autosomal alleles with dominance. His example is based on some data which include several children from each of a number of families, rather than just one child per family. By contrast, the present paper (like the work of MacCluer and Schull) is concerned solely with the statistically simpler situation in which there is no deliberate selection of two or more trios having the same mother and putative father, i.e., no deliberate selection of more than one child from a family (see assumption 5 below).

Received June 14, 1965.

Supported in part by U. S. Public Health Service research grants GM 09358 and GM 12868 from the National Institutes of Health.

MacCluer and Schull consider the estimation of λ for three different cases: two autosomal alleles without dominance, two autosomal alleles with dominance, and two sex-linked alleles with dominance. For each case, they designate a special grouping of the data and then estimate λ by maximizing the likelihood function which applies after this grouping has already been effected. Because of the predesignated grouping, the resulting estimates (which could be considered to be maximum-likelihood estimates in a special context) do not utilize the entire information available. With our approach in the present paper, we obtain maximum-likelihood estimates based on the *ungrouped* data and thereby make optimal use of the data.

OUTLINE OF THE MAXIMUM-LIKELIHOOD METHOD

Before considering specific cases, we begin with a general discussion of the uses of the maximum-likelihood method in estimating λ, the proportion of nonpaternity. Suppose the genetic trait we are dealing with manifests t different possible phenotypes; thus, e.g., $t = 3$ for the case of two alleles without dominance and $t = 2$ for the case of two alleles with dominance. We may arbitrarily identify the t phenotypes by the numbers from 1 to t. Thus, for the case of the M-N blood system, we may assign the numbers 1, 2, and 3 to the phenotypes M, MN, and N respectively. Let x_{ijk} denote the observed number of trios for which the putative father has phenotype i, the mother has phenotype j, and the child has phenotype k. Thus x_{311} would denote the number of trios falling into the category in which the putative father is N, the mother M, and the child M. Note that the sum of all the x_{ijk}'s will be n, the total number of trios in the sample.

Next we define f_{ijk} to be the expected frequency of trios in which the putative father is of phenotype i, the mother is j, and the child k. The t^3 different f_{ijk}'s will add up to 1. Some f_{ijk}'s will automatically be zero, as will the corresponding x_{ijk}'s. In general, each f_{ijk} will be a function of λ and of the gene frequencies. The various f_{ijk}'s for some of the more important cases are listed in Tables 1 to 4.

We may assume that the x_{ijk}'s follow a multinomial distribution, so that their likelihood function is

$$L = \frac{n!}{\prod\limits_{i,j,k} x_{ijk}!} \prod_{i,j,k} f_{ijk}{}^{x_{ijk}} \qquad (1)$$

The logarithm of L is equal to a constant (i.e., a term not dependent on λ or the gene frequencies) plus

$$L^* = \sum_{i,j,k} x_{ijk} \log f_{ijk} \qquad (2)$$

The products in (1) and the summation in (2) are taken over all (i,j,k) combinations except those for which f_{ijk} is automatically zero.

To apply the method of maximum likelihood, we maximize L (1), or, equivalently and more conveniently, we may maximize $L^*(2)$. Specifically, we proceed as follows. Suppose there are just two alleles, for which the respective gene frequencies are p and q $(= 1 - p)$, where p is unknown. Considering $L^*(2)$ as a function of λ and p, we take its partial derivatives with respect to λ and p, and then solve the equations

$$\frac{\partial L^*}{\partial \lambda} = 0 \tag{3a}$$

$$\frac{\partial L^*}{\partial p} = 0 \tag{3b}$$

for λ and p. The resulting solutions, which may be denoted by $\hat{\lambda}$ and $\hat{p}$, are the maximum-likelihood estimates of λ and p respectively. The generalized Newton-Raphson method may be used to solve the system (3a,3b), which is a system of two equations in two unknowns.

Suppose there are three alleles, with corresponding (unknown) gene frequencies p, q, and r $(= 1 - p - q)$. Then we solve a system of three equations in three unknowns and obtain the estimates $\hat{\lambda}$, $\hat{p}$, and $\hat{q}$; the system is similar to (3a,3b) except that there is a third equation involving the partial derivative of L^* with respect to q.

The maximum-likelihood estimates will be approximately normally distributed if n is sufficiently large. It would appear that, for a given n, the normal approximation to the distribution of $\hat{\lambda}$ will not be as good when λ is close to zero as it is when λ is somewhat larger, because of greater skewness when λ is near zero. The approximate variance of $\hat{\lambda}$, which we will call $s^2(\hat{\lambda})$, may be calculated. If, e.g., there are two alleles and we are estimating p as well as λ, then we first compute the elements of the 2 × 2 matrix

$$\underset{\sim}{I} = \begin{bmatrix} \sum\limits_{i,j,k} f_{ijk} \left(\dfrac{\partial \log f_{ijk}}{\partial \lambda} \right)^2 & \sum\limits_{i,j,k} f_{ijk} \left(\dfrac{\partial \log f_{ijk}}{\partial \lambda} \right) \left(\dfrac{\partial \log f_{ijk}}{\partial p} \right) \\ \sum\limits_{i,j,k} f_{ijk} \left(\dfrac{\partial \log f_{ijk}}{\partial \lambda} \right) \left(\dfrac{\partial \log f_{ijk}}{\partial p} \right) & \sum\limits_{i,j,k} f_{ijk} \left(\dfrac{\partial \log f_{ijk}}{\partial p} \right)^2 \end{bmatrix} \tag{4}$$

using the estimated values $\hat{\lambda}$ and $\hat{p}$ for λ and p respectively. To obtain $s^2(\hat{\lambda})$, we calculate the element in the upper left-hand corner of $\underset{\sim}{I}^{-1}$ and divide it by n.

Assumptions

One matter which we have so far ignored is that of the assumptions upon which the above method rests. For our assumptions, we follow much the same ones which were used (either explicitly or implicitly) by MacCluer and Schull and by Li. They are as follows:

1. We assume that the phenotypes of all of the $3n$ individuals contained in the sample have been correctly ascertained and that no errors in diagnosis have been made.

2. For every trio we assume that the "mother" is indeed the true mother of the child. In other words, we assume that all mothers are correctly identified and that the frequency of nonmaternity is zero. Although questions of nonmaternity might be of interest in certain situations, the present paper makes no attempt to consider such problems.

3. We assume that the Hardy-Weinberg equilibruim conditions hold. Furthermore, with respect to the mother, putative father, and true father (if different from the putative father) pertaining to any child, we assume that, for the trait under examination, the genotypes of these three (or two) individuals are determined randomly and mutually independently. This set of assumptions concerning Hardy-Weinberg equilibrium and random choice of partners would probably not be satisfied in any population comprised of different sociological groups whose respective gene frequencies are not all equal.

4. We assume that the expected proportion of nonpaternity is equal to the same value λ for all combinations of genotypes of mother and putative father.

5. We assume that the n trios in the sample are selected *independently* of each other (and randomly). This means that the investigator should avoid a sampling scheme which deliberately selects more than one child from a family. Perhaps the most severe adverse effect of systematically using more than one child from a family would be with respect to $s^2(\hat{\lambda})$ rather than $\hat{\lambda}$ itself. It would appear that, if the n trios in the sample represent far fewer than n different families, then the estimator $s^2(\hat{\lambda})$ could seriously underestimate the true variance of $\hat{\lambda}$, due to statistical complications caused by family-to-family variation in the parameter represented by λ. MacCluer and Schull evidently assume independent selection of the n trios. Although Li makes no such assumption (and, in fact, exhibits an example where the sampling is by families), neither does he make any attempt to estimate the variance of $\hat{\lambda}$.

Further Uses of the Maximum-Likelihood Technique

Provided again that the five assumptions which we just discussed are satisfied, our estimation of λ can be effected via the maximum-likelihood technique under circumstances other than those for which the basic description of the method was given above. We now indicate several such situations:

1. Instead of being unknown, the gene frequencies for the population may sometimes be rather precisely known as a result of previous studies. In such circumstances, what we do is to substitute these known values of the gene frequencies into equation (3a) and then solve this single equation for λ.

The solution, which we again call $\hat{\lambda}$, is our estimate of λ. Its approximate variance $s^2(\hat{\lambda})$ is $(1/n)$ times the reciprocal of the element in the upper left-hand corner of $\underset{\sim}{I}$ (4), where this element is evaluated by using the known values for the gene frequencies and using the estimate $\hat{\lambda}$ for λ. It is apparent that the application of the maximum-likelihood method becomes somewhat simpler if the gene frequencies are known, because now we need to solve only a single equation (3a) rather than a system of two or more equations such as the system (3a,3b).

2. Even when the gene frequencies are really unknown, one might still like to use a scheme similar to the one just described in order to reduce the computational burden. One might be inclined, e.g., to estimate the gene frequencies on the basis of the phenotypes of the mothers and putative fathers, and then substitute these estimates into (3a) and solve the resulting equation for λ in order to get an estimate $\hat{\lambda}$. Although it would be favored with definite computational advantages, such a $\hat{\lambda}$ obviously would not be the exact maximum-likelihood estimate but would appear to provide an amply close approximation if n is moderately large.*

3. The $3n$ individuals in the sample might sometimes be examined with respect to more than just one genetic trait. We would expect that the extra information provided by an additional trait would lead to an estimate of λ with reduced variance. If we can assume independence among the two or more traits with respect to the distribution of their genotypes in the population, then the f_{ijk} formulas are easily obtained. The maximum-likelihood method can thus still be applied but becomes more complicated than before. However, these complications will be minimized if all the gene frequencies are known (refer to paragraph 1 just above) or if the scheme just described in paragraph 2 is used, since then only a single lengthy equation in λ will have to be solved. An example below will show how to obtain t and the f_{ijk} formulas when there is more than one trait.

4. Instead of the individual x_{ijk}'s, we may sometimes only have available a tally of the trios with respect to a coarser and more condensed classification, and it may not be possible to recover the x_{ijk}'s if the original data are no longer accessible. Thus, the trios having an M putative father, N mother, and MN child might have been combined with the trios having an N putative father, M mother, and MN child, so that the sum $(x_{132} + x_{312})$ is available but the individual values x_{132} and x_{312} are not. In such situations the maximum-likelihood method can still be applied if we work with the ex-

*If a $\hat{\lambda}$ is obtained in the fashion described in this paragraph, the question will arise as to what to use for $s^2(\hat{\lambda})$, the estimated variance of $\hat{\lambda}$. One possibility would be to obtain $s^2(\hat{\lambda})$ as in the preceding paragraph (except, of course, with the estimated instead of the known gene frequencies); this $s^2(\hat{\lambda})$ would have computational advantages but apparently would tend to underestimate the true variance of $\hat{\lambda}$. A second possibility, which would probably tend to give more accurate results, would be to calculate $s^2(\hat{\lambda})$ along lines similar to those indicated in the discussion accompanying equation (4).

pected frequencies which pertain to the coarser classification. Thus, the expected frequency for the category just described would be $(f_{132} + f_{312})$, which we would find (via Table 1) to be equal to $p^2q^2(2 - \lambda)$. By maximizing the likelihood function which pertains to the coarser classification, we obtain an estimate of λ which is the maximum-likelihood estimate based on the limited available data but which obviously is not the same as the maximum-likelihood estimate which would have resulted from the complete data consisting of the x_{ijk}'s.

5. Even when all the x_{ijk}'s are available, one could still deliberately group the data in some fashion and then obtain an estimate of λ by maximizing the likelihood function which pertains to the resulting coarser classification. Such a procedure would not utilize all the available data and would thus generally produce an estimator with greater variance; a second difficulty would be that one would have to make a decision as to how to do the grouping. However, there might be offsetting advantages (such as reduced computational requirements, in particular). The deliberate grouping of the categories of trios could be effected in many different ways, one of which is utilized by MacCluer and Schull.

APPLICATIONS TO DIFFERENT CASES

Now that we have completed our general discussion of the estimation of λ via the method of maximum likelihood, we may turn to the more specific matter of applications. The first thing that will be needed in any application will be the formulas for the f_{ijk}'s. Tables 1 through 4 present these formulas for the following four cases respectively: two autosomal alleles without dominance (as exemplified by the MN blood system), two autosomal alleles (denoted by C and c) with dominance, two sex-linked alleles (denoted by G and g) with dominance, and the ABO blood system when four phenotypes (A, B, AB, and O) are distinguished. For the first three tables there are two gene frequencies, p and q $(= 1 - p)$, which are the respective frequencies of M and N in Table 1, of C and c in Table 2, and of G and g in Table 3. In Table 4, the respective gene frequencies for A, B, and O are p, q, and r $(= 1 - p - q)$. At the top of each table are given the identifications for the t possible phenotypes (note that $t = 3$ for Table 1, $t = 2$ for Tables 2 and 3, and $t = 4$ for Table 4). For each of the t^3 possible categories of trios (some of which are null), the body of each table lists the formulas for f_{ijk}. Except in Table 4, each f_{ijk} formula is first presented as the product of three elements, which are respectively the expected frequency of the putative father's phenotype, the expected frequency of the mother's phenotype, and the expected frequency of the child's phenotype given the phenotypes of the putative father and the mother. Thus, Table 2 tells us that the expected frequency for the category of trios having a c putative father, C mother, and C child is

$$f_{211} = (q^2)(p^2 + 2pq)\left\{\frac{1}{q+1}(1 - \lambda) + \frac{pq+1}{q+1}\lambda\right\}$$

TABLE 1. FORMULAS FOR THE f_{ijk}'S FOR THE CASE OF TWO AUTOSOMAL ALLELES WITHOUT DOMINANCE (AS EXEMPLIFIED BY THE MN BLOOD SYSTEM) AND DATA (x_{ijk}'S) FOR A NUMERICAL EXAMPLE

Gene frequencies: p = frequency of M, q = frequency of N.
Phenotype identifications: 1 refers to M, 2 refers to MN. 3 refers to N.
Formula for D/λ: $D/\lambda = pq(1-pq)$.

Phenotype of				
Putative father	Mother	Child	Expected frequency for category	Data of Gershowitz
M	M	M	$f_{111} = (p^2)(p^2)\{(1-\lambda)+p\lambda\}$	$x_{111} = 14$
M	M	MN	$f_{112} = (p^2)(p^2)\{q\lambda\}$	$x_{112} = 5$
M	MN	M	$f_{121} = (p^2)(2pq)\{\frac{1}{2}(1-\lambda)+\frac{1}{2}p\lambda\}$	$x_{121} = 13$
M	MN	MN	$f_{122} = (p^2)(2pq)\{\frac{1}{2}\}$	$x_{122} = 16$
M	MN	N	$f_{123} = (p^2)(2pq)\{\frac{1}{2}q\lambda\}$	$x_{123} = 0$
M	N	MN	$f_{132} = (p^2)(q^2)\{(1-\lambda)+p\lambda\}$	$x_{132} = 20$
M	N	N	$f_{133} = (p^2)(q^2)\{q\lambda\}$	$x_{133} = 1$
MN	M	M	$f_{211} = (2pq)(p^2)\{\frac{1}{2}(1-\lambda)+p\lambda\}$	$x_{211} = 13$
MN	M	MN	$f_{212} = (2pq)(p^2)\{\frac{1}{2}(1-\lambda)+q\lambda\}$	$x_{212} = 9$
MN	MN	M	$f_{221} = (2pq)(2pq)\{\frac{1}{4}(1-\lambda)+\frac{1}{2}p\lambda\}$	$x_{221} = 20$
MN	MN	MN	$f_{222} = (2pq)(2pq)\{\frac{1}{2}\}$	$x_{222} = 41$
MN	MN	N	$f_{223} = (2pq)(2pq)\{\frac{1}{4}(1-\lambda)+\frac{1}{2}q\lambda\}$	$x_{223} = 21$
MN	N	MN	$f_{232} = (2pq)(q^2)\{\frac{1}{2}(1-\lambda)+p\lambda\}$	$x_{232} = 15$
MN	N	N	$f_{233} = (2pq)(q^2)\{\frac{1}{2}(1-\lambda)+q\lambda\}$	$x_{233} = 19$
N	M	M	$f_{311} = (q^2)(p^2)\{p\lambda\}$	$x_{311} = 2$
N	M	MN	$f_{312} = (q^2)(p^2)\{(1-\lambda)+q\lambda\}$	$x_{312} = 14$
N	MN	M	$f_{321} = (q^2)(2pq)\{\frac{1}{2}p\lambda\}$	$x_{321} = 1$
N	MN	MN	$f_{322} = (q^2)(2pq)\{\frac{1}{2}\}$	$x_{322} = 14$
N	MN	N	$f_{323} = (q^2)(2pq)\{\frac{1}{2}(1-\lambda)+\frac{1}{2}q\lambda\}$	$x_{323} = 16$
N	N	MN	$f_{332} = (q^2)(q^2)\{p\lambda\}$	$x_{332} = 1$
N	N	N	$f_{333} = (q^2)(q^2)\{(1-\lambda)+q\lambda\}$	$x_{333} = 10$
Any	M	N	$f_{113} = f_{213} = f_{313} = 0$	
Any	N	M	$f_{131} = f_{231} = f_{331} = 0$	
Totals			$\sum_i\sum_j\sum_k f_{ijk} = 1$	$n = 265$

where q^2 is the expected frequency of a c putative father, $(p^2 + 2pq)$ is the expected frequency of a C mother, $1/(q+1)$ is the probability of a C child if the putative father (who is c) is the true father, and $(pq+1)/(q+1)$ is the expected frequency of a C child if the putative father is not the true father. The formula for f_{211} simplifies to $pq^2(1+pq\lambda)$.

The f_{ijk} formulas in Table 3 pertain to the situation where the only examined trios are those in which the child is a girl. In the case of a trait determined by sex-linked alleles, data on trios with male children will provide no information about λ, since a son receives a Y chromosome from his father no matter who his father is. Thus better estimation of λ will be achieved if the available resources are expended only on the examination of trios with female children, in which event Table 3 will be fully applicable. In case trios with male children have also been examined, however, then the situation may be a little less simple; in some circumstances we might simply discard these trios and still utilize Table 3, but in other circumstances

TABLE 2. FORMULAS FOR THE f_{ijk}'S FOR THE CASE OF TWO AUTOSOMAL ALLELES WITH DOMINANCE

Gene frequencies: p = frequency of C, q = frequency of c.
Phenotype identifications: 1 refers to C, 2 refers to c.
Formula for D/λ: D /λ = pq^4.

Phenotype of			
Putative father	Mother	Child	Expected frequency for category
C	C	C	$f_{111} = (p^2 + 2pq)\ (p^2 + 2pq)\ \left\{\dfrac{2q+1}{(q+1)^2}(1-\lambda) + \dfrac{pq+1}{q+1}\lambda\right\} = p^2\ (2q + 1 - q^3\lambda)$
C	C	c	$f_{112} = (p^2 + 2pq)(p^2 + 2pq)\ \{[q^2/(q+1)^2](1-\lambda) + [q^2/(q+1)]\lambda\}$ $= p^2q^2(1 + q\lambda)$
C	c	C	$f_{121} = (p^2 + 2pq)(q^2)\ \{[1/(q+1)](1-\lambda) + p\lambda\} = pq^2(1 - q\lambda)$
C	c	c	$f_{122} = (p^2 + 2pq)(q^2)\ \{[q/(q+1)](1-\lambda) + q\lambda\} = pq^3(1 + q\lambda)$
c	C	C	$f_{211} = (q^2)(p^2 + 2pq)\ \{[1/(q+1)](1-\lambda) + [(pq+1)/(q+1)]\lambda\}$ $= pq^2(1 + pq\lambda)$
c	C	c	$f_{212} = (q^2)(p^2 + 2pq)\ \{[q/(q+1)](1-\lambda) + [q^2/(q+1)]\lambda\}$ $= pq^3(1 - p\lambda)$
c	c	C	$f_{221} = (q^2)(q^2)\ \{p\lambda\} = pq^4\lambda$
c	c	c	$f_{222} = (q^2)(q^2)\ \{(1-\lambda) + q\lambda\} = q^4(1 - p\lambda)$

TABLE 3. FORMULAS FOR THE f_{ijk}'S FOR THE CASE OF TWO SEX-LINKED ALLELES WITH DOMINANCE WHEN THE CHILD IN EACH TRIO IS A GIRL

Gene frequencies: p = frequency of G, q = frequency of g.
Phenotype identifications: 1 refers to G, 2 refers to g.
Formula for D/λ: $D/\lambda = q^2(1-q^2)$.

Phenotype of			
Putative father	Mother	Daughter	Expected frequency for category
G	G	G	$f_{111} = (p)(p^2 + 2pq)\ \{(1-\lambda) + [(pq+1)/(q+1)]\lambda\} = p^2(q + 1 - q^2\lambda)$
G	G	g	$f_{112} = (p)(p^2 + 2pq)\ \{[q^2/(q+1)]\lambda\} = p^2q^2\lambda$
G	g	G	$f_{121} = (p)(q^2)\ \{(1-\lambda) + p\lambda\} = pq^2(1 - q\lambda)$
G	g	g	$f_{122} = (p)(q^2)\ \{q\lambda\} = pq^3\lambda$
g	G	G	$f_{211} = (q)(p^2 + 2pq)\ \{[1/(q+1)](1-\lambda) + [(pq+1)/(q+1)]\lambda\}$ $= pq(1 + pq\lambda)$
g	G	g	$f_{212} = (q)(p^2 + 2pq)\ \{[q/(q+1)](1-\lambda) + [q^2/(q+1)]\lambda\}$ $= pq^2(1 - p\lambda)$
g	g	G	$f_{221} = (q)(q^2)\ \{p\lambda\} = pq^3\lambda$
g	g	g	$f_{222} = (q)(q^2)\ \{(1-\lambda) + q\lambda\} = q^3(1 - p\lambda)$

more complicated techniques might be called for (in order to take advantage, e.g., of the information about gene frequencies which these trios provide).

Following MacCluer and Schull, we use D to denote the expected frequency of trios which constitute detectable instances of nonpaternity (i.e., instances in which it is genetically impossible for the putative father to be the true father). Then the quantity D/λ is the fraction of nonpaternity which is detectable. At the top of each of the four tables we indicate the applicable formula for D/λ, because this formula can be utilized to advantage to obtain a preliminary value for the estimate of λ. The D/λ formulas of the first three tables were already presented by MacCluer and Schull.

TABLE 4. FORMULAS FOR THE f_{ijk}'S FOR THE CASE OF THE ABO BLOOD SYSTEM WHEN FOUR PHENOTYPES ARE DISTINGUISHED

Gene frequencies: p = frequency of A, q = frequency of B, r = frequency of O.
Phenotype identifications: 1 refers to A, 2 refers to B, 3 refers to AB, 4 refers to O.
Formula for D/λ: $D/\lambda = pq(p + 2r)(p + r)^2 + pq(q + 2r)(q + r)^2 + 4pqr^2 + 2pqr^3 + r^4(p + q)$.

Phenotype of			
Putative father	Mother	Child	Expected frequency of category
A	A	A	$f_{111} = p^2(p + r)(p + 3r)(1 - \lambda) + p^2(p + 2r)(p^2 + 3pr + r^2)\lambda$
A	A	B	$f_{112} = p^2qr(p + 2r)\lambda$
A	A	AB	$f_{113} = p^2q(p + 2r)(p + r)\lambda$
A	A	O	$f_{114} = p^2r^2(1 - \lambda) + p^2r^2(p + 2r)\lambda$
A	B	A	$f_{121} = pqr(p + r)(1 - \lambda) + p^2qr(p + 2r)\lambda$
A	B	B	$f_{122} = pqr(q + r)(1 - \lambda) + pq(p + 2r)(q^2 + 3qr + r^2)\lambda$
A	B	AB	$f_{123} = pq(p + r)(q + r)(1 - \lambda) + p^2q(p + 2r)(q + r)\lambda$
A	B	O	$f_{124} = pqr^2(1 - \lambda) + pqr^2(p + 2r)\lambda$
A	AB	A	$f_{131} = p^2q(p + 2r)(1 - \lambda) + p^2q(p + 2r)(p + r)\lambda$
A	AB	B	$f_{132} = p^2qr(1 - \lambda) + p^2q(p + 2r)(q + r)\lambda$
A	AB	AB	$f_{133} = p^2q(p + r)(1 - \lambda) + p^2q(p + 2r)(p + q)\lambda$
A	O	A	$f_{141} = pr^2(p + r)(1 - \lambda) + p^2r^2(p + 2r)\lambda$
A	O	B	$f_{142} = pqr^2(p + 2r)\lambda$
A	O	O	$f_{144} = pr^3(1 - \lambda) + pr^3(p + 2r)\lambda$
B	A	A	$f_{211} = pqr(p + r)(1 - \lambda) + pq(q + 2r)(p^2 + 3pr + r^2)\lambda$
B	A	B	$f_{212} = pqr(q + r)(1 - \lambda) + pq^2r(q + 2r)\lambda$
B	A	AB	$f_{213} = pq(p + r)(q + r)(1 - \lambda) + pq^2(q + 2r)(p + r)\lambda$
B	A	O	$f_{214} = pqr^2(1 - \lambda) + pqr^2(q + 2r)\lambda$
B	B	A	$f_{221} = pq^2r(q + 2r)\lambda$
B	B	B	$f_{222} = q^2(q + r)(q + 3r)(1 - \lambda) + q^2(q + 2r)(q^2 + 3qr + r^2)\lambda$
B	B	AB	$f_{223} = pq^2(q + 2r)(q + r)\lambda$
B	B	O	$f_{224} = q^2r^2(1 - \lambda) + q^2r^2(q + 2r)\lambda$
B	AB	A	$f_{231} = pq^2r(1 - \lambda) + pq^2(q + 2r)(p + r)\lambda$
B	AB	B	$f_{232} = pq^2(q + 2r)(1 - \lambda) + pq^2(q + 2r)(q + r)\lambda$
B	AB	AB	$f_{233} = pq^2(q + r)(1 - \lambda) + pq^2(q + 2r)(p + q)\lambda$
B	O	A	$f_{241} = pqr^2\ (q + 2r)\lambda$
B	O	B	$f_{242} = qr^2(q + r)(1 - \lambda) + q^2r^2(q + 2r)\lambda$
B	O	O	$f_{244} = qr^3(1 - \lambda) + qr^3(q + 2r)\lambda$
AB	A	A	$f_{311} = p^2q(p + 2r)(1 - \lambda) + 2p^2q(p^2 + 3pr + r^2)\lambda$
AB	A	B	$f_{312} = p^2qr(1 - \lambda) + 2p^2q^2r\lambda$
AB	A	AB	$f_{313} = p^2q(p + r)(1 - \lambda) + 2p^2q^2(p + r)\lambda$
AB	A	O	$f_{314} = 2p^2qr^2\lambda$
AB	B	A	$f_{321} = pq^2r(1 - \lambda) + 2p^2q^2r\lambda$
AB	B	B	$f_{322} = pq^2(q + 2r)(1 - \lambda) + 2pq^2(q^2 + 3qr + r^2)\lambda$
AB	B	AB	$f_{323} = pq^2(q + r)(1 - \lambda) + 2p^2q^2(q + r)\lambda$
AB	B	O	$f_{324} = 2pq^2r^2\lambda$
AB	AB	A	$f_{331} = p^2q^2(1 - \lambda) + 2p^2q^2(p + r)\lambda$
AB	AB	B	$f_{332} = p^2q^2(1 - \lambda) + 2p^2q^2(q + r)\lambda$
AB	AB	AB	$f_{333} = 2p^2q^2(1 - \lambda) + 2p^2q^2(p + q)\lambda$
AB	O	A	$f_{341} = pqr^2(1 - \lambda) + 2p^2qr^2\lambda$
AB	O	B	$f_{342} = pqr^2(1 - \lambda) + 2pq^2r^2\lambda$
AB	O	O	$f_{344} = 2pqr^3\lambda$
O	A	A	$f_{411} = pr^2(p + r)(1 - \lambda) + pr^2(p^2 + 3pr + r^2)\lambda$
O	A	B	$f_{412} = pqr^3\lambda$

TABLE 4. (CONTINUED)

Phenotype of			
Putative father	Mother	Child	Expected frequency of category
O	A	AB	$f_{413} = pqr^2(p+r)\lambda$
O	A	O	$f_{414} = pr^3(1-\lambda) + pr^4\lambda$
O	B	A	$f_{421} = pqr^3\lambda$
O	B	B	$f_{422} = qr^2(q+r)(1-\lambda) + qr^2(q^2+3qr+r^2)\lambda$
O	B	AB	$f_{423} = pqr^2(q+r)\lambda$
O	B	O	$f_{424} = qr^3(1-\lambda) + qr^4\lambda$
O	AB	A	$f_{431} = pqr^2(1-\lambda) + pqr^2(p+r)\lambda$
O	AB	B	$f_{432} = pqr^2(1-\lambda) + pqr^2(q+r)\lambda$
O	AB	AB	$f_{433} = pqr^2(p+q)\lambda$
O	O	A	$f_{441} = pr^4\lambda$
O	O	B	$f_{442} = qr^4\lambda$
O	O	O	$f_{444} = r^4(1-\lambda) + r^5\lambda$
Any	AB	O	$f_{134} = f_{234} = f_{334} = f_{434} = 0$
Any	O	AB	$f_{143} = f_{243} = f_{343} = f_{443} = 0$

Tables 1 through 4 can be utilized in the situation where the individuals in the sample have been examined with respect to more than one trait. An example should serve to make this clear. Suppose we are dealing with two traits, one involving two autosomal alleles (M and N) without dominance and the other involving two autosomal alleles (C and c) with dominance. In this situation, there are $3 \times 2 = 6$ possible phenotype combinations, and so we consider t to be equal to 6. We need a formula for the expected frequency of each of the $t^3 = 216$ categories of trios. As an example, we consider just one of these categories, the one for which the putative father of the trio has phenotypes N and C, the mother is M and c, and the child is MN and C. Then the expected frequency for this category is

$$\{q^2(P^2+2PQ)\}\,(p^2Q^2)\left\{\frac{1}{Q+1}(1-\lambda) + qP\lambda\right\}$$

where p and q are the respective gene frequencies for M and N, and P and Q (in place of p and q) denote the respective gene frequencies for C and c. The above formula can be obtained very simply if we just look at f_{312} in Table 1 and f_{121} (first formula) in Table 2: $q^2(P^2 + 2PQ)$ is the product of the first element of f_{312} by the first element of f_{121}, p^2Q^2 is the product of the two second elements, $1/(Q + 1)$ is the product of the two coefficients of $(1 - \lambda)$, and qP is the product of the two coefficients of λ.

The last column of Table 1 displays some unpublished data of Gershowitz concerning $n = 265$ trios, comprised wholly of Detroit Negroes, which were examined with respect to the MN blood system. The distribution of these 265 observed trios into the various categories is exhibited in the table. At the end of the paper, we will use these data to work out a numerical example. The 265 trios were obtained independently and represent 265 separate families or 795 (3 × 265) distinct individuals; hence, our assumption 5 is certainly tenable.

Gershowitz's 265 trios in Table 1 include the 243 trios which were previously reported by MacCluer and Schull (p. 200) in connection with their numerical example.

The Maximum-Likelihood Equations

The system of maximum-likelihood equations is the system (3a,3b), or sometimes there may be just one equation or three or more equations rather than two. Although nothing more than elementary differential calculus is involved in obtaining the equations (3a,3b), it may nevertheless serve to clarify the maximum-likelihood estimation procedure if we present the explicit equations for two of the most common cases. The two cases we will look at will be the ones covered by Tables 1 and 2.

For a trait involving two autosomal alleles without dominance, we use Table 1 and find that the system of maximum-likelihood equations (3a,3b) is

$$\frac{x_1(p-\tfrac{1}{2})}{p\lambda+\tfrac{1}{2}(1-\lambda)}+\frac{x_2(q-\tfrac{1}{2})}{q\lambda+\tfrac{1}{2}(1-\lambda)}-\frac{x_3q}{p\lambda+(1-\lambda)}-\frac{x_4p}{q\lambda+(1-\lambda)}+\frac{x_D}{\lambda}=0 \tag{5a}$$

$$\frac{x_1\lambda}{p\lambda+\tfrac{1}{2}(1-\lambda)}-\frac{x_2\lambda}{q\lambda+\tfrac{1}{2}(1-\lambda)}+\frac{x_3\lambda}{p\lambda+(1-\lambda)}-\frac{x_4\lambda}{q\lambda+(1-\lambda)}+\frac{x_M}{p}-\frac{x_N}{q}=0 \tag{5b}$$

where we define

$$x_1 = x_{211} + x_{221} + x_{232} \tag{6a}$$

$$x_2 = x_{212} + x_{223} + x_{233} \tag{6b}$$

$$x_3 = x_{111} + x_{121} + x_{132} \tag{6c}$$

$$x_4 = x_{312} + x_{323} + x_{333} \tag{6d}$$

$$x_D = x_{112} + x_{123} + x_{133} + x_{311} + x_{321} + x_{332} \tag{6e}$$

$$\begin{aligned} x_M = {} & 4x_{111} + 4x_{112} + 3x_{121} + 3x_{122} + 3x_{123} + 2x_{132} + 2x_{133} \\ & + 3x_{211} + 3x_{212} + 2x_{221} + 2x_{222} + 2x_{223} + x_{232} + x_{233} \\ & + 3x_{311} + 2x_{312} + 2x_{321} + x_{322} + x_{323} + x_{332} \end{aligned} \tag{6f}$$

and

$$\begin{aligned} x_N = {} & x_{112} + x_{121} + x_{122} + 2x_{123} + 2x_{132} + 3x_{133} \\ & + x_{211} + x_{212} + 2x_{221} + 2x_{222} + 2x_{223} + 3x_{232} + 3x_{233} \\ & + 2x_{311} + 2x_{312} + 3x_{321} + 3x_{322} + 3x_{323} + 4x_{332} + 4x_{333} \end{aligned} \tag{6g}$$

Equations (5a,5b) are solved simultaneously for λ and p. It may be of interest to note that x_M(6f) is the number of M genes among all putative fathers and all mothers plus the small number of children's M genes which could not possibly have been received either from the mother or the putative father; x_N(6g) has a similar interpretation with respect to N genes. Note also that x_D(6e) is the number of trios which represent detectable instances of nonpaternity.

In the event that the gene frequencies p and q are known, we substitute these known values into equation (5a). Then we discard equation (5b) altogether and just solve the single equation (5a) for λ.

For the case of a trait involving two autosomal alleles with dominance, we use Table 2 and find that the maximum-likelihood equations (3a,3b) are

$$-\frac{x_{111}q^3}{2q+1-q^3\lambda}+\frac{(x_{112}+x_{122})q}{1+q\lambda}-\frac{x_{121}q^2}{1-q^2\lambda}+\frac{x_{211}pq}{1+pq\lambda}-\frac{(x_{212}+x_{222})p}{1-p\lambda}+\frac{x_{221}}{\lambda}=0 \tag{7a}$$

and

$$\frac{x_{111}(3q^2\lambda-2)}{2q+1-q^3\lambda}-\frac{(x_{112}+x_{122})\lambda}{1+q\lambda}+\frac{2x_{121}q\lambda}{1-q^2\lambda}+\frac{x_{211}(1-2p)\lambda}{1+pq\lambda}-\frac{(x_{212}+x_{222})\lambda}{1-p\lambda}+\frac{x_C}{p}-\frac{x_c}{q}=0 \tag{7b}$$

where we define

$$x_C = 2x_{111}+2x_{112}+x_{121}+x_{122}+x_{211}+x_{212}+x_{221}$$

and

$$x_c = 2x_{112}+2x_{121}+3x_{122}+2x_{211}+3x_{212}+4x_{221}+4x_{222}$$

We solve equations (7a) and (7b) simultaneously for λ and p; or, if p is known, we solve (7a) for λ.

Numerical Example

By using the data of Gershowitz which are exhibited in the last column of Table 1, we now present a numerical example to illustrate the estimation of λ by the maximum-likelihood method. To start our calculations, we obtain the seven numbers (6a–6g). The first one (6a) is $x_1 = 13 + 20 + 15 = 48$. We also find $x_2 = 49$, $x_3 = 47$, $x_4 = 40$, $x_D = 10$, $x_M = 536$, and $x_N = 534$.

Our next step is to substitute these seven numbers into (5a,5b), and then solve this equation system (5a,5b) for λ and p. To obtain the solution, we may employ the generalized Newton-Raphson method, which is an iterative procedure.

In order to utilize this iterative method, we first need preliminary values for λ and p which can be used in the initial iteration and which (preferably) will tend to be fairly close to the solution values $\hat{\lambda}$ and $\hat{p}$. We might use $x_M/(x_M + x_N) = 536/(536 + 534) = .5009$ as our value of p for the initial iteration. To obtain a preliminary value for λ, we could proceed as follows. Looking at the top of Table 1, we find that $D/\lambda = pq(1 - pq)$. If p is .5009, the numerical value of D/λ is then .1875. As the value of λ for the initial iteration, we can use $(x_D/n)/.1875 = (10/265)/.1875 = .2013$.

The generalized Newton-Raphson method is a standard technique which is adequately discussed in textbooks on numerical analysis, and MacCluer and Schull (pp. 200–201) also have exhibited a particular example of its use. For these reasons, it is perhaps best to save space and omit all of the detailed computations associated with the Newton-Raphson procedure (except for the material in the preceding paragraph, which demonstrates one way of calculating the initial values). We present simply the final result of the computations, which is that

$$\hat{\lambda} = .2062, \qquad \hat{p} = .5012$$

constitute the solution of (5a,5b) and hence constitute the maximum-likelihood estimates of λ and p.

To obtain $s^2(\hat{\lambda})$, the estimated variance of the estimate $\hat{\lambda}$, we first need the matrix $\underset{\sim}{I}(4)$. The calculation of $\underset{\sim}{I}$ involves numerical evaluation of the f_{ijk}'s of Table 1 and of the first derivatives of their logarithms, using .2062 for λ and .5012 for p. Again we omit the cumbersome details and present just the result, which is

$$\underset{\sim}{I} = \begin{bmatrix} 1.01384 & -.00152 \\ -.00152 & 16.2362 \end{bmatrix}$$

Now we find

$$s^2(\hat{\lambda}) = \frac{1}{265} \times \frac{16.2362}{(1.01384)(16.2362) - (-.00152)^2}$$
$$= \frac{1}{265} \times \frac{16.2362}{16.4609} = .003722$$

In case we are also interested in the estimated variance of $\hat{p}$, we write

$$s^2(\hat{p}) = \frac{1}{265} \times \frac{1.01384}{16.4609} = .0002324$$

When n is large enough that the normal approximation to the distribution of $\hat{\lambda}$ is reasonably accurate, the formula

$$\hat{\lambda} - 1.96\sqrt{s^2(\hat{\lambda})} \leq \lambda \leq \hat{\lambda} + 1.96\sqrt{s^2(\hat{\lambda})} \qquad (8)$$

provides a 95% confidence interval for λ. For the present example, (8) becomes

$$.0866 \leq \lambda \leq .3258$$

However, the normal approximation probably is still not as accurate as we might like it to be, even in this situation with an n of 265.

Although all the above results were obtained with a desk calculator, the use of a computer might have reduced the labor. For obtaining the maximum-likelihood estimate of λ in more complicated situations, a computer would almost be a necessity. It is always possible to estimate λ by a means which requires virtually no calculation, such as by the procedure which we utilized earlier which consisted simply of dividing (x_D/n) by an approximate value of (D/λ). The variance of a simple estimator such as this may not be too much larger than the variance of the maximum-likelihood estimator in some situations (among which our example appears to be included) but may be markedly larger in other situations.

As we indicated earlier, 243 of the 265 trios of Table 1 above are included in the tabulation which MacCluer and Schull made of Gershowitz's data and which they utilized for their numerical example (pp. 200–201). Professor Gershowitz has informed us that the estimation procedure of MacCluer and Schull has been rerun using the data as it is in Table 1 (rather than the former tabulation), and he has kindly sent us a copy of these recalculations. The results of the estimation procedure of MacCluer and Schull, correct to four significant figures, are as follows: $\hat{\lambda} = .2105$, $\hat{p} = .5192$, $s^2(\hat{\lambda}) = .003907$, $s^2(\hat{p}) = .0004611$. The confidence interval for λ of the form (8) thus turns out to be $.0880 \leq \lambda \leq .3330$.

Thus both $s^2(\hat{\lambda})$ and $s^2(\hat{p})$ are larger with the MacCluer-Schull procedure than with our maximum-likelihood procedure. This is not surprising in view of the asymptotic efficiency properties of maximum-likelihood estimates. However, the improvement in $s^2(\hat{\lambda})$ which the maximum-likelihood procedure produces over the MacCluer-Schull procedure is not especially large in this particular example. If p had not been so close to ½, it would appear that the contrast would have been somewhat greater, because more information about λ would have been extractable from the categories with an MN putative father and thus a larger contribution to the element in the upper left-hand corner of $\underset{\sim}{I}(4)$ would have resulted.

SUMMARY

This paper attempts to clarify the use of the maximum-likelihood method for estimating the proportion (λ) of nonpaternity or extramarital illegitimacy. The phenotypes of the putative father, the mother, and the child in

each of a number of trios constitute the basis for classifying each trio into just one of many categories. Each category is uniform as to the combination of the three phenotypes; i.e., dissimilar categories are not merged. The expected frequency for each category will generally depend on λ, and λ is estimated by utilizing the expected frequency and the observed number of trios for each category. The paper concludes with a numerical example.

ACKNOWLEDGMENT

The authors are indebted to Professor Henry Gershowitz of the University of Michigan for his generous efforts in providing us with the detailed breakdown of his data in the form in which it appears in Table 1 and for furnishing us with certain recalculations based on the data.

REFERENCES

Cramér, H. 1946. *Mathematical Methods of Statistics*. Princeton, New Jersey: Princeton University Press.

Fraser, D. A. S. 1958. *Statistics—An Introduction*. New York: Wiley.

Li, C. C. 1961. *Human Genetics*. New York: McGraw-Hill.

MacCluer, J. W., and Schull, W. J. 1963. On the estimation of the frequency of nonpaternity. *Amer. J. Hum. Genet.* 15: 191–202.