The Annals of Statistics
1974, Vol. 2, No. 5, 911-924

LOG-LINEAR MODELS FOR FREQUENCY TABLES DERIVED BY INDIRECT OBSERVATION: MAXIMUM LIKELIHOOD EQUATIONS[1]

BY SHELBY J. HABERMAN

University of Chicago

Frequency tables are examined in which some cells are not distinguishable. Log-linear models are proposed for these tables which lead to likelihood equations closely related to those associated with log-linear models for conventional frequency tables. Just as in conventional tables, the maximum likelihood equations are shown to be the same under Poisson or multinomial sampling. Applications are made to the problem of estimation of gene frequencies from observed phenotype frequencies.

1. Introduction. In many statistical applications, a frequency table can only be observed indirectly. The statistician is then faced with the task of describing the behavior of the original table by use of the limited available data. Specifically, it may be the case that a table $\mathbf{n} = \{n_i : i \in I\}$ of frequencies indexed by a finite, nonempty index set I is of interest, but the available data consist of a table $\mathbf{n}^* = \{n_j^* : j \in J\}$ such that

$$I = \bigcup_{j \in J} J_j \tag{1.1}$$

for some disjoint sets J_j, $j \in J$, and

$$n_j^* = \sum_{i \in J_j} n_i \,. \tag{1.2}$$

Given $\mathbf{n}^*$, one may wish to estimate the mean $\mathbf{m} = \{m_i\}$ of $\mathbf{n}$ under the assumption that $\mathbf{n}$ satisfies a log-linear model in which $m_i > 0$ for $i \in I$ and $\boldsymbol{\mu} = \{\log m_i\} \in \mathscr{M}$, a linear manifold in R^I, the space of real I-tuples. This estimate may be used to provide a test of the validity of the model that $\boldsymbol{\mu} \in \mathscr{M}$ or to estimate various unknown parameters.

Estimation problems of this type arise in a variety of applications. In this paper, primary emphasis will be given to problems in genetics in which gene probabilities in a population are estimated on the basis of phenotype frequencies. For instance, in the *ABO* blood group, a pair of genes with alleles *O*, *A*, and *B* determine the phenotypes *O*, *A*, *B*, and *AB*. Since *O* is recessive with respect to *A* and *B*, phenotype *O* corresponds to the genotype *OO*, *A* corresponds to *AA* or *AO*, *B* corresponds to *BB* or *BO*, and *AB* corresponds to *AB*. One can take a

Received September 1972; revised September 1973.

[1] This research was supported in part by National Science Foundation Grant NSF GP-323037x and in part by the Department of Statistics, University of Chicago. The section on genetic theory has benefited from a number of conversations with Mr. John Wiorkowski.

AMS 1970 *subject classifications.* Primary 62F10; Secondary 62P10.

Key words and phrases. Log-linear models, contingency tables, maximum likelihood estimation, genetic models, indirect observation.

random sample of N individuals and observe the number n_j^* of individuals with phenotype $j \in J = \{O, A, B, AB\}$. Given these data, one may wish to estimate the probabilities p_O, p_A, and p_B that genes in the population have alleles O, A, and B. This estimation problem is a classic one in genetics (see Elandt-Johnson (1971, page 395)). To show that it is an example of the class of problems considered in this paper, let $n_{ii'}$ be the number of individuals in the sample such that the gene from the father has allele i and the gene from the mother has allele i'. Then $I = \{O, A, B\} \times \{O, A, B\}$, $J_O = \{OO\}$, $J_A = \{AO, OA, AA\}$, $J_B = \{BO, OB, BB\}$, and $J_{AB} = \{AB, BA\}$. If the population is in Hardy-Weinberg equilibrium, then the expected value $m_{ii'}$ of $n_{ii'}$ is $Np_ip_{i'}$. If $p_i > 0$ for $i \in \{O, A, B\}$, then this equilibrium hypothesis is readily seen to be equivalent to the hypothesis that $\mu_{ii'} = \log m_{ii'}$ satisfies

$$\mu_{ii'} = \alpha + \beta_i + \beta_{i'}, \qquad (i, i') \in I, \tag{1.3}$$

for some α and $\{\beta_i\}$ such that $\sum \beta_i = 0$. In other words, the equilibrium hypothesis is equivalent to the hypothesis that $\boldsymbol{\mu} \in \mathcal{M}$, where $\mathcal{M}$ is the set of $\boldsymbol{\mu}$ which satisfy (1.3). Since $\mathbf{m}$ is determined by p_O, p_A, and p_B and in turn determines these probabilities, estimation of $\mathbf{m}$ is equivalent to estimation of p_O, p_A, and p_B. Thus this type of problem illustrates one example of the sampling situation considered in this paper.

Other examples arise in biological assay and social psychology. In biological assay, a logit model for quantal response with natural mortality involves a log-linear model in which the cause of death cannot be directly observed. In social psychology, studies of group behavior such as those analyzed by Fienberg and Larntz (1971) attempt to describe behavior of a group in terms of unobserved actions of individuals in the group. The general sampling problems considered in this paper also have been explored by Chen (1972) and Haberman (1971).

The method used in this paper to estimate the mean vector $\mathbf{m}$ is that of maximum likelihood. To obtain maximum likelihood estimates, we first consider maximum likelihood estimation under the assumption that underlying observations n_i, $i \in I$, are independent Poisson random variables. These results are then applied to sampling situations in which the vector $\mathbf{n}$ consists of one or more independent multinomial samples satisfying the same regularity conditions as those in Haberman (1970, 1973, and 1974). The basic results of Sections 2 and 3 are that the maximum likelihood equations are the same for any of these sampling schemes and may be expressed in terms of conditional and unconditional expected values. In Section 4, these maximum likelihood equations are computed for some problems involving estimation of gene frequencies by observation of phenotype frequencies. Sections 5 and 6 are devoted to problems of existence and uniqueness of solutions of the maximum likelihood equations. It is shown that solutions need not be unique, but that some solution must exist whenever a Fisher-consistent estimate of $\boldsymbol{\mu}$ is available and $n_j^* > 0$ for $j \in J$.

2. Maximum likelihood equations for Poisson samples. In this section, it is

assumed that the underlying frequencies n_i, $i \in I$, are independent Poisson random variables with means $m_i = e^{\mu_i}$, so that the observed frequencies n_j^*, $j \in J$, are independent Poisson random variables with means

$$E(n_j^*) = \sum_{i \in J_j} e^{\mu_i} .$$

Thus the log-likelihood $l(\mathbf{n}^*, \boldsymbol{\mu})$ satisfies

$$l(\mathbf{n}^*, \boldsymbol{\mu}) = \sum_{j \in J} [n_j^* \log (\sum_{i \in J_j} e^{\mu_i}) - \sum_{i \in J_j} e^{\mu_i} - \log n_j^*!] \tag{2.1}$$

for all $\boldsymbol{\mu} \in \mathscr{M}$.

The principal result of this section relates conditional expected values, unconditional expected values, and maximum likelihood estimates. Let $\mathbf{m}(\boldsymbol{\mu}) = \{\exp \mu_i\}$ be the unconditional expected value of $\mathbf{n}$, and let $m(\boldsymbol{\mu} \mid \mathbf{n}^*)$ be the conditional expected value of $\mathbf{n}$ given $\mathbf{n}^*$. Note that

$$\mathbf{m}(\boldsymbol{\mu} \mid \mathbf{n}^*) = \{n_j^* e^{\mu_i} / \sum_{i' \in J_j} e^{\mu_{i'}} : i \in J_j, j \in J\} . \tag{2.2}$$

Let $P_{\mathscr{M}}$ be the orthogonal projection on $\mathscr{M}$ with respect to the usual inner product $(\cdot, \cdot)$ defined for $\mathbf{x}$ and $\mathbf{y}$ in R^I by

$$(\mathbf{x}, \mathbf{y}) = \sum_{i \in I} x_i y_i .$$

Then the following theorem holds:

THEOREM 1. *If $l(\mathbf{n}^*, \boldsymbol{\mu})$ has a maximum at $\hat{\boldsymbol{\mu}}$, then*

$$P_{\mathscr{M}} \mathbf{m}(\hat{\boldsymbol{\mu}} \mid \mathbf{n}^*) = P_{\mathscr{M}} \mathbf{m}(\hat{\boldsymbol{\mu}}) . \tag{2.3}$$

If the vectors $\boldsymbol{\mu}^{(t)}$, $1 \leqq t \leqq s$, span $\mathscr{M}$, then (2.3) is equivalent to the equation

$$(\boldsymbol{\mu}^{(t)}, \mathbf{m}(\hat{\boldsymbol{\mu}} \mid \mathbf{n}^*)) = (\boldsymbol{\mu}^{(t)}, \mathbf{m}(\hat{\boldsymbol{\mu}})) , \qquad 1 \leqq t \leqq s . \tag{2.4}$$

PROOF. Consider the first differential $dl_{\mu}(\mathbf{n}^*, \boldsymbol{\nu})$ of $l(\mathbf{n}^*, \boldsymbol{\mu})$. This differential is a linear function on $\mathscr{M}$ such that for $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ in $\mathscr{M}$,

$$l(\mathbf{n}^*, \boldsymbol{\mu} + \boldsymbol{\nu}) = l(\mathbf{n}^*, \boldsymbol{\mu}) + dl_{\mu}(\mathbf{n}^*, \boldsymbol{\nu}) + \mathscr{O}(\boldsymbol{\nu}) , \tag{2.5}$$

where $\mathscr{O}(\boldsymbol{\nu})/||\boldsymbol{\nu}|| \to 0$ as $||\boldsymbol{\nu}|| \to 0$. Here $||\boldsymbol{\nu}||$, the norm of $\boldsymbol{\nu}$, is defined to be $(\boldsymbol{\nu}, \boldsymbol{\nu})^{\frac{1}{2}}$. Since

$$\begin{aligned} &n_j^* \log (\sum_{i \in J_j} e^{\mu_i + \nu_i}) - \sum_{i \in J_j} e^{\mu_i + \nu_i} \\ &\quad = n_j^* \log (\sum_{i \in J_j} e^{\mu_i}) + n_j^* (\sum_{i \in J_j} e^{\mu_i} \nu_i) / (\sum_{i \in J_j} e^{\mu_i}) \\ &\qquad - \sum_{i \in J_j} e^{\mu_i} - \sum_{i \in J_j} e^{\mu_i} \nu_i + \mathscr{O}(\boldsymbol{\nu}) , \end{aligned} \tag{2.6}$$

where $\mathscr{O}(\boldsymbol{\nu})/||\boldsymbol{\nu}|| \to 0$ as $||\boldsymbol{\nu}|| \to 0$, (2.5) and the definitions of $\mathbf{m}(\boldsymbol{\mu})$, $\mathbf{m}(\boldsymbol{\mu} \mid \mathbf{n}^*)$, and $(\cdot, \cdot)$ imply that

$$\begin{aligned} dl_{\mu}(\mathbf{n}^*, \boldsymbol{\nu}) &= \sum_{j \in J} n_j^* (\sum_{i \in J_j} e^{\mu_i} \nu_i) / (\sum_{i \in J_j} e^{\mu_i}) - \sum_{i \in I} e^{\mu_i} \nu_i \\ &= (\mathbf{m}(\boldsymbol{\mu} \mid \mathbf{n}^*) - \mathbf{m}(\boldsymbol{\mu}), \boldsymbol{\nu}) . \end{aligned} \tag{2.7}$$

If the log likelihood has a maximum at $\hat{\boldsymbol{\mu}}$, then $dl_{\hat{\mu}}(\mathbf{n}^*, \boldsymbol{\nu}) = 0$ for each $\boldsymbol{\nu} \in \mathscr{M}$. This condition is equivalent to the condition

$$P_{\mathscr{M}} \mathbf{m}(\hat{\boldsymbol{\mu}} \mid \mathbf{n}^*) = P_{\mathscr{M}} \mathbf{m}(\hat{\boldsymbol{\mu}}) . \tag{2.8}$$

Alternatively, if the span of the vectors $\boldsymbol{\mu}^{(t)}$, $1 \leqq t \leqq s$, is $\mathscr{M}$, then (2.8) is equivalent to the condition

$$(\boldsymbol{\mu}^{(t)}, \mathbf{m}(\hat{\boldsymbol{\mu}} \mid \mathbf{n}^*)) = (\boldsymbol{\mu}^{(t)}, \mathbf{m}(\hat{\boldsymbol{\mu}})), \qquad 1 \leqq t \leqq s. \tag{2.9}$$

Thus the maximum likelihood equations reduce to the requirement that certain linear combinations of conditional and unconditional expected values must agree. □

In the trivial case when $I = J$, the observed and unobserved tables coincide and $\mathbf{m}(\boldsymbol{\mu} \mid \mathbf{n}^*) = \mathbf{n}$. Equations (2.8) and (2.9) then coincide with the corresponding equations for log-linear models derived by Haberman (1970, 1973, and 1974). Thus in the case of independent Poisson observations, the maximum likelihood equations are generalizations for the corresponding equations when $\mathbf{n}$ is observed directly.

3. Maximum likelihood equations for multinomial samples. Comparable results may also be obtained if the underlying frequencies $\mathbf{n}$ consist of independent vectors $\{n_i : i \in I_k\}$, $1 \leqq k \leqq r$, $r \geqq 1$, with multinomial distributions. It is assumed that the I_k, $1 \leqq k \leqq r$, are disjoint sets with union I. The vector $\{n_i : i \in I_k\}$ is assumed to be a sample of size $N_k > 0$ and to have expected value $\{m_i(\boldsymbol{\mu}) : i \in I_k\}$. It is assumed that the vectors $\boldsymbol{\nu}^{(k)}$, $1 \leqq k \leqq r$, are in $\mathscr{M}$, where

$$\begin{aligned} \nu_i^{(k)} &= 1, & i \in I_k, \\ &= 0, & i \notin I_k, \end{aligned}$$

and each J_j, $j \in J$, is contained in some I_k, $1 \leqq k \leqq r$. The condition that $\boldsymbol{\nu}^{(k)} \in \mathscr{M}$, $1 \leqq k \leqq r$, corresponds to the condition used by Haberman (1970, 1973, and 1974) to examine multinomial samples. The condition that each J_j is in some I_k is simply the requirement that the sample from which an observation is derived must be known.

To derive the maximum likelihood equations for this sampling situation, we follow Haberman (1970, 1973, and 1974), and note that the constraint that

$$\textstyle\sum_{i \in I_k} m_i(\boldsymbol{\mu}) = N_k, \qquad 1 \leqq k \leqq r, \tag{3.1}$$

insures that $\mathbf{m}(\boldsymbol{\mu})$ is uniquely determined by $\boldsymbol{\mu}' = P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu}$, where $\mathscr{M} - \mathscr{N}$ is the orthogonal complement of $\mathscr{N}$ relative to $\mathscr{M}$ and $\mathscr{N}$ is the span of the vectors $\boldsymbol{\nu}^{(k)}$, $1 \leqq k \leqq r$. An element $m_i(\boldsymbol{\mu})$ such that $i \in I_k$ then satisfies the equation

$$m_i(\boldsymbol{\mu}) = N_k e^{\mu_i} / \textstyle\sum_{i' \in I_k} e^{\mu_{i'}}. \tag{3.2}$$

Conversely, if $\boldsymbol{\mu}' \in \mathscr{M} - \mathscr{N}$ and (3.2) holds, then $\boldsymbol{\mu} \in \mathscr{M}$, (3.1) holds, and $\boldsymbol{\mu}' = P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu}$.

If $A_k = \{j \in J : J_j \subset I_k\}$, then the vectors $\{n_j^* : j \in A_k\}$, $1 \leqq k \leqq r$, are independent multinomial random vectors with respective sample sizes N_k and means

$$\{\textstyle\sum_{i \in J_j} m_i(\boldsymbol{\mu}) : j \in A_k\}.$$

Given (3.2), the log likelihood $l^{(m)}(\mathbf{n}^*, \boldsymbol{\mu}')$ can be written

$$l^{(m)}(\mathbf{n}^*, \boldsymbol{\mu}') = \sum_{k=1}^{r} \{\sum_{j \in A_k} n_j^* \log (\sum_{i \in J_j} e^{\mu_i'}) - N_k \log (\sum_{i \in I_k} e^{\mu_i'}) + \log N_k! - \sum_{j \in A_k} \log n_j^*!\} . \tag{3.3}$$

Corresponding to Theorem 1, we have the following parallel result:

THEOREM 2. *If $l^{(m)}(\mathbf{n}^*, \boldsymbol{\mu}')$ has a maximum at $\hat{\boldsymbol{\mu}} = P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu}$, where $\hat{\boldsymbol{\mu}}$ satisfies* (3.1), *then*

$$P_{\mathscr{M}}\mathbf{m}(\hat{\boldsymbol{\mu}} \mid \mathbf{n}^*) = P_{\mathscr{M}}\mathbf{m}(\hat{\boldsymbol{\mu}}) . \tag{3.4}$$

PROOF. Proceeding as in the case of independent Poisson observations, we find that the differential $dl_{\boldsymbol{\mu}'}^{(m)}(\mathbf{n}^*, \boldsymbol{\nu})$ for $\boldsymbol{\nu} \in \mathscr{M} - \mathscr{N}$ satisfies

$$\begin{aligned} dl_{\boldsymbol{\mu}'}^{(m)}(\mathbf{n}^*, \boldsymbol{\nu}) &= \sum_{k=1}^{r} \{\sum_{j \in A_k} n_j^* (\sum_{i \in J_j} e^{\mu_i'} \nu_i)/(\sum_{i \in J_j} e^{\mu_i'}) \\ &\qquad - N_k (\sum_{i \in I_k} e^{\mu_i'} \nu_i)/(\sum_{i \in I_k} e^{\mu_i'})\} \\ &= (\mathbf{m}(\boldsymbol{\mu} \mid \mathbf{n}^*) - \mathbf{m}(\boldsymbol{\mu}), \boldsymbol{\nu}) , \end{aligned} \tag{3.5}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$ are related by (3.2). Thus if a maximum of $l^{(m)}(\mathbf{n}^*, \boldsymbol{\mu}')$ is achieved at $\hat{\boldsymbol{\mu}}'$, then for the corresponding vector $\hat{\boldsymbol{\mu}} \in \mathscr{M}$,

$$P_{\mathscr{M}-\mathscr{N}}\mathbf{m}(\hat{\boldsymbol{\mu}} \mid \mathbf{n}^*) = P_{\mathscr{M}-\mathscr{N}}\mathbf{m}(\hat{\boldsymbol{\mu}}) \cdot \tag{3.6}$$

Since by (2.2),

$$\begin{aligned} (\mathbf{m}(\hat{\boldsymbol{\mu}} \mid \mathbf{n}^*), \boldsymbol{\nu}^{(k)}) &= \sum_{i \in I_k} m_i(\hat{\boldsymbol{\mu}} \mid \mathbf{n}^*) = \sum_{j \in A_k} \sum_{i \in J_j} m_i(\hat{\boldsymbol{\mu}} \mid \mathbf{n}^*) \\ &= \sum_{j \in A_k} n_j^* = N_k = \sum_{i \in I_k} m_i(\hat{\boldsymbol{\mu}}) \\ &= (\mathbf{m}(\hat{\boldsymbol{\mu}}), \boldsymbol{\nu}^{(k)}) , \qquad 1 \leqq k \leqq r , \end{aligned} \tag{3.7}$$

it follows from the definition of $\mathscr{N}$ that

$$P_{\mathscr{N}}\mathbf{m}(\hat{\boldsymbol{\mu}} \mid \mathbf{n}^*) = P_{\mathscr{N}}\mathbf{m}(\hat{\boldsymbol{\mu}}) . \tag{3.8}$$

Thus $\hat{\boldsymbol{\mu}}$ is a solution of (3.6) if and only if

$$P_{\mathscr{M}}\mathbf{m}(\hat{\boldsymbol{\mu}} \mid \mathbf{n}^*) = P_{\mathscr{M}}\mathbf{m}(\hat{\boldsymbol{\mu}}) . \tag{3.9}$$

□

The equation for $\hat{\boldsymbol{\mu}}$ coincides with (2.3) and is equivalent to (2.4). Thus the maximum likelihood equations for the two sampling methods are equivalent. Hence if $\hat{\boldsymbol{\mu}}$ satisfies (3.1) and $P_{\mathscr{M}-\mathscr{N}}\hat{\boldsymbol{\mu}}$ is a maximum of $l^{(m)}(\mathbf{n}^*, \boldsymbol{\mu}')$, then $\hat{\boldsymbol{\mu}}$ satisfies the maximum likelihood equation (2.3) for the maximum of $l(\mathbf{n}^*, \boldsymbol{\mu})$ for $\boldsymbol{\mu} \in \mathscr{M}$. Conversely, if $\hat{\boldsymbol{\mu}}$ is a maximum of $l(\mathbf{n}^*, \boldsymbol{\mu})$ for $\boldsymbol{\mu} \in \mathscr{M}$, the $\hat{\boldsymbol{\mu}}$ satisfies (3.9). Since $\mathscr{N} \subset \mathscr{M}$, $\hat{\boldsymbol{\mu}}$ satisfies (3.8) and (3.1). Therefore, $\hat{\boldsymbol{\mu}}$ satisfies the maximum likelihood equation (3.6) and the constraint (3.1). In fact, since $l(\mathbf{n}^*, \boldsymbol{\mu})$ and $l^{(m)}(\mathbf{n}^*, P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu})$ differ only by a constant for any $\boldsymbol{\mu}$ satisfying (3.1), $P_{\mathscr{M}-\mathscr{N}}\hat{\boldsymbol{\mu}}$ must be a maximum of $l^{(m)}(\mathbf{n}^*, \boldsymbol{\mu}')$ for $\boldsymbol{\mu}' \in \mathscr{M} - \mathscr{N}$.

4. Applications to genetics. In this section, $h \geqq 1$ pairs of genes determine the phenotype $j \in J$ of an individual in a given population. Each gene pair, or locus, g, $g = 1, \cdots, h$, has alleles in a set A_g of at least two elements, so that

an individual has the combination of alleles, or genotype, $\mathbf{i} = \{[i(g), i'(g)]: g = 1, \cdots, h\}$, where for gene pair $g = 1, \cdots, h$, $i(g) \in A_g$ is the allele from the father and $i'(g) \in A_g$ is the allele from the mother. This distinction between genes from the father and from the mother is not generally made in genetic literature, but it is useful in the development of the formulas used in the section. The probability that a randomly chosen individual from the population has allele combination $\mathbf{i}$ is assumed to be $\pi_\mathbf{i}$, $I = \{\mathbf{i}: \pi_\mathbf{i} > 0\}$, and for each $j \in J$, J_j is the set of genotype $\mathbf{i} \in I$ which result in phenotype j. On the basis of genetic theory, $\mathbf{i} \in J_j$ implies $\mathbf{d} \in J_j$ if for each g, $g = 1, \cdots, h$, either $i(g) = d'(g)$ and $i'(g) = d(g)$ or $i(g) = d(g)$ and $i'(g) = d'(g)$.

If a random sample of N individuals is observed such that $n_j{}^*$ subjects have phenotype $j \in J$ and n_i subjects have genotype $\mathbf{i} \in I$, then the assumptions of Section 3 are satisfied whenever $\{\log \pi_\mathbf{i}\} \in \mathscr{M}$, a linear manifold containing the unit vector $\mathbf{e} = \{1: \mathbf{i} \in I\}$. This assertion follows since, in the notation of Section 3, $r = 1$, $I_1 = I$, $\boldsymbol{\nu}^{(1)} = \mathbf{e} \in \mathscr{M}$, $J_j \subset I$ for each $j \in J$, and

$$\boldsymbol{\mu} = \{\log N\pi_\mathbf{i}\} = (\log N)\mathbf{e} + \{\log \pi_\mathbf{i}\} \in \mathscr{M}.$$

Since numerous common genetic models for the probability vector $\boldsymbol{\pi} = \{\pi_\mathbf{i}\}$ correspond to log-linear models for $\boldsymbol{\mu}$ or $\{\log \pi_\mathbf{i}\}$, (3.4) may be used to construct maximum likelihood equations for many models in genetics. The following subsections illustrate this procedure.

4.1. *The case of one pair of alleles.* In this subsection, it is assumed that one pair of alleles determines the phenotype. The population is assumed to be in Hardy–Weinberg equilibrium, and the distribution of alleles from the father is assumed to be the same as the distribution from the mother. These assumptions hold under random mating, provided there are no differences of fitness of different genotypes or other forms of selection (see Cepellini, Siniscalco, and Smith (1955)).

Under these assumptions, the probability that a randomly selected individual has allele $i \in A$ from the father and allele $i' \in A$ from the mother is

$$\pi_{ii'} = p_i p_{i'}, \tag{4.1}$$

where p_i is the probability that a gene has allele $i \in A$. Here subscripts corresponding to the index g have been deleted since g is always 1. If $p_i > 0$ for each $i \in A$, then (4.1) is equivalent to the assumption that for some α and $\{\beta_i: i \in A\}$,

$$\mu_{ii'} = \log (N\pi_{ii'}) = \alpha + \beta_i + \beta_{i'}, \qquad (i, i') \in I = A \times A, \tag{4.2}$$

where

$$\textstyle\sum_{i \in A} \beta_i = 0. \tag{4.3}$$

The set $\mathscr{M}$ of $\boldsymbol{\mu} \in R^I$ such that for some α and $\{\beta_i: i \in A\}$, (4.2) and (4.3) are satisfied, is a linear manifold which is spanned by the vectors $\boldsymbol{\mu}^{(t)}$, $t \in A$, where $\mu_{ii'}^{(t)} = \delta_{it} + \delta_{i't}$ and δ_{it} is the Kronecker δ function.

Given the spanning vectors $\boldsymbol{\mu}^{(t)}$, $t \in A$, (2.4) reduces to the conditions

$$\begin{aligned}\sum_{i\in A}\sum_{i'\in A} m_{ii'}(\hat{\boldsymbol{\mu}}\,|\,\mathbf{n}^*)(\delta_{it}+\delta_{i't}) &= m_{t+}(\hat{\boldsymbol{\mu}}\,|\,\mathbf{n}^*)+m_{+t}(\boldsymbol{\mu}\,|\,\mathbf{n}^*)\\ &= m_{t+}(\hat{\boldsymbol{\mu}})+m_{+t}(\hat{\boldsymbol{\mu}})\\ &= 2N\hat{p}_t\,, \quad t\in A\,.\end{aligned}$$

Since for any J_j, $j \in J$, genetic theory requires that $(i, i') \in J_j$ implies $(i', i) \in J_j$, it follows that if $(i, i') \in J_j$, then

$$\begin{aligned}m_{ii'}(\mu\,|\,n^*) &= n_j^*\hat{p}_i\hat{p}_{i'}/\textstyle\sum_{(d,d')\in J_j}\hat{p}_d\hat{p}_{d'}\\ &= m_{i'i}(\hat{\boldsymbol{\mu}}\,|\,\mathbf{n}^*)\,.\end{aligned}$$

Thus $m_{t+}(\hat{\boldsymbol{\mu}}\,|\,\mathbf{n}^*) = m_{+t}(\hat{\boldsymbol{\mu}}\,|\,\mathbf{n}^*)$ and

$$\begin{aligned}\hat{p}_t &= \frac{1}{N}m_{t+}(\boldsymbol{\mu}\,|\,\mathbf{n}^*)\\ &= \frac{1}{N}\textstyle\sum_{j\in J} n_j^*[\sum_{(t,i')\in J_j}\hat{p}_t\hat{p}_{i'}/\sum_{(i,i')\in J_j}\hat{p}_i\hat{p}_{i'}]\,, \qquad t\in A\,.\end{aligned} \tag{4.4}$$

The left-hand side is the maximum likelihood estimate of the gene frequency p_t. The right-hand side is the expected proportion of genes in the population with allele t, given the observations $\mathbf{n}^*$. This equation has been used by Ceppellini, Siniscalco, and Smith (1955) as the basis of the gene counting method of computing maximum likelihood estimates for problems of the type considered in this section.

A classical example of these maximum likelihood equations is associated with the *ABO* blood group. In this example, the maximum likelihood equations are

$$\hat{p}_O = \frac{1}{N}[n_O^* + n_A^*\hat{p}_O/(\hat{p}_A+2\hat{p}_O) + n_B^*\hat{p}_O/(\hat{p}_B+2\hat{p}_O)]\,, \tag{4.5}$$

$$\hat{p}_A = \frac{1}{N}[n_A^*(\hat{p}_O+\hat{p}_A)/(\hat{p}_A+2\hat{p}_O) + \tfrac{1}{2}n_{AB}^*]\,, \tag{4.6}$$

and

$$\hat{p}_B = \frac{1}{N}[n_B^*(\hat{p}_O+\hat{p}_B)/(\hat{p}_B+2\hat{p}_O) + \tfrac{1}{2}n_{AB}^*] \tag{4.7}$$

(see Elandt-Johnson (1971, page 400)).

To illustrate derivation of these equations, consider (4.5). The right-hand side is the expected proportion of genes with allele O among the $2N$ genes in the sample, given the observed values of n_O^*, n_A^*, n_B^*, and n_{AB}^*. Given that a subject has phenotype O, there must be 2 genes for the subject with allele O. Thus $2n_O^*$ genes with allele O must be present among the n_O^* subjects with phenotype O. If a subject has phenotype A, then the possible pairs of alleles are AO, OA, and AA. The estimated expected number of alleles O is then

$$\frac{\hat{p}_A\hat{p}_O+\hat{p}_O\hat{p}_A}{\hat{p}_A\hat{p}_O+\hat{p}_O\hat{p}_A+\hat{p}_A^2} = \frac{2\hat{p}_O}{2\hat{p}_O+\hat{p}_A}\,.$$

Thus the estimated expected number of alleles O corresponding to the $n_A{}^*$ subjects with phenotype A is $2n_A{}^*\hat{p}_O/(2\hat{p}_O + \hat{p}_A)$. Similar arguments show that the estimated expected number of alleles O corresponding to the $n_B{}^*$ subjects with phenotype B is $2n_B{}^*\hat{p}_O/(2\hat{p}_O + \hat{p}_B)$. If one adds the expected values corresponding to these phenotype frequencies, and divides by $2N$, one obtains (4.5).

Both Ceppellini, Siniscalco, and Smith (1955) and Elandt-Johnson (1971) also apply gene counting to the *MNSs* blood group. Further material on this method is given by Yasuda and Kimura (1968) and Haberman (1971).

4.2. *The case of several pairs of alleles where linkage is not present.* If h pairs of alleles determine the phenotype, no linkage is present, no other selection factors exist, and the distribution of alleles for fathers and for mothers is the same, then the probability $\pi_{\mathbf{i}}$ that a randomly chosen member of the population has alleles $\mathbf{i} = \{[i(g), i'(g)]: g = 1, \cdots, h\}$ satisfies

$$\pi_{\mathbf{i}} = \prod_{g=1}^{h} p_{i(g)}^{(g)} p_{i'(g)}^{(g)}, \qquad \mathbf{i} \in \prod_{g=1}^{h} (A_g \times A_g), \tag{4.8}$$

where $p_i{}^{(g)} > 0$ is the probability that a gene at locus g has allele $i \in A_g$. Here (4.8) is equivalent to the hypothesis that for some α and $\{\beta_i{}^{(g)}: i \in A_g\}$, $g = 1, \cdots, h$,

$$\mu_{\mathbf{i}} = \alpha + \sum_{g=1}^{h} (\beta_{i(g)}^{(g)} + \beta_{i'(g)}^{(g)}) \tag{4.9}$$

and

$$\sum_{i \in A_g} \beta_i{}^{(g)} = 0, \qquad g = 1, \cdots, h. \tag{4.10}$$

The linear manifold $\mathscr{M}$ of $\boldsymbol{\mu}$ such that (4.9) and (4.10) hold for some α and $\{\beta_i{}^{(g)}\}$, $g = 1, \cdots, h$, is spanned by the vectors $\boldsymbol{\mu}^{(t,g)}$, $t \in A_g$, $g = 1, \cdots, h$, where

$$\mu_{\mathbf{i}}{}^{(t,g)} = \delta_{i(g)t} + \delta_{i'(g)t}.$$

Instead of (4.4), one now has

$$\hat{p}_t{}^{(g)} = \frac{1}{N} m_t{}^{(g)}(\hat{\boldsymbol{\mu}} \mid \mathbf{n}^*), \qquad t \in A_g, \; g = 1, \cdots, h, \tag{4.11}$$

where

$$m_t{}^{(g)}(\boldsymbol{\mu} \mid \mathbf{n}^*) = \sum_{j \in J} n_j{}^* \left[\frac{\sum_{\mathbf{i} \in J_j, i(g)=t} \prod_{g'=1} \hat{p}_{i(g')} \hat{p}_{i'(g')}}{\sum_{\mathbf{i} \in J_j} \prod_{g'=1}^{h} p_{i(g')} p_{i'(g')}} \right]. \tag{4.12}$$

In (4.11), the left-hand side is the estimated proportion of genes in the population at locus g with allele t, and the right-hand side is the expected proportion of genes in the sample at locus g with allele t. Therefore, (4.11) is a direct generalization of (4.4).

To illustrate application of gene counting in the multiple loci case, consider the two-loci model of Koler, Jones, Wasi, and Pootrukul (1971). In this model, two unlinked loci determine the phenomenon of decreased synthesis of human hemoglobin α-chains. In this model, $J = \{1, 2, 3, 4, 5\}$ and the possible phenotypes, with corresponding values of j in parenthesis, are normal (1), silent carrier (2), α-thalassemia trait (3), Hb H disease (4), and Hydrops foetalis (5). The first

locus has alleles T and t', while the second locus has alleles T and t. Thus $h = 2$, $A_1 = \{T, t'\}$, and $A_2 = \{T, t\}$. The probabilities $p_T{}^{(1)}$, $p_{t'}^{(1)}$, $p_T{}^{(2)}$, and $p_t{}^{(2)}$ are to be estimated.

The correspondence between genotypes and phenotypes is such that J_1 contains $\{(T, T), (T, T)\}$, $\{(T, T), (T, t)\}$, $\{(T, T), (t, T)\}$, and $\{(T, T), (t, t)\}$, J_2 contains $\{(T, t'), (T, T)\}$ and $\{(t', T), (T, T)\}$, J_3 contains $\{(T, t'), (T, t)\}$, $\{(t', T), (T, t)\}$, $\{(T, t'), (t, T)\}$, $\{(t', T), (t, T)\}$, $\{(T, t'), (t, t)\}$, $\{(t', T), (t, t)\}$, and $\{(t', t'), (T, T)\}$, J_4 contains $\{(t', t'), (T, t)\}$ and $\{(t', t'), (t, T)\}$ and J_5 contains $\{(t', t'), (t, t)\}$. The maximum likelihood equations are then

$$\hat{p}_T{}^{(1)} = \frac{1}{N}[n_1{}^* + \tfrac{1}{2}n_2{}^* + n_3{}^*(2\hat{p}_T{}^{(1)}\hat{p}_T{}^{(2)}\hat{p}_t{}^{(2)} + \hat{p}_T{}^{(1)}\hat{p}_t{}^{(2)}\hat{p}_t{}^{(2)}) \div (4\hat{p}_{t'}^{(1)}\hat{p}_T{}^{(2)}\hat{p}_t{}^{(2)} + 2\hat{p}_T{}^{(1)}\hat{p}_t{}^{(2)}\hat{p}_t{}^{(2)} + \hat{p}_{t'}^{(1)}\hat{p}_T{}^{(2)}\hat{p}_T{}^{(2)})],$$

$$\hat{p}_{t'}^{(1)} = \frac{1}{N}[\tfrac{1}{2}n_2{}^* + n_3{}^*(2\hat{p}_T{}^{(1)}\hat{p}_T{}^{(2)}\hat{p}_t{}^{(2)} + \hat{p}_T{}^{(1)}\hat{p}_t{}^{(2)}\hat{p}_t{}^{(2)} + \hat{p}_{t'}^{(1)}\hat{p}_T{}^{(2)}\hat{p}_T{}^{(2)}) \div (4\hat{p}_{t'}^{(1)}\hat{p}_T{}^{(2)}\hat{p}_t{}^{(2)} + 2\hat{p}_T{}^{(1)}\hat{p}_t{}^{(2)}\hat{p}_t{}^{(2)} + \hat{p}_{t'}^{(1)}\hat{p}_T{}^{(2)}\hat{p}_T{}^{(2)}) + n_4{}^* + n_5{}^*],$$

$$\hat{p}_T{}^{(2)} = \frac{1}{N}[n_1{}^*\hat{p}_T{}^{(2)} + n_2{}^* + n_3{}^*(2\hat{p}_T{}^{(1)}\hat{p}_T{}^{(2)}\hat{p}_t{}^{(2)} + \hat{p}_{t'}^{(1)}\hat{p}_T{}^{(2)}\hat{p}_T{}^{(2)}) \div (4\hat{p}_{t'}^{(1)}\hat{p}_T{}^{(2)}\hat{p}_t{}^{(2)} + 2\hat{p}_T{}^{(1)}\hat{p}_t{}^{(2)}\hat{p}_t{}^{(2)} + \hat{p}_{t'}^{(1)}\hat{p}_T{}^{(2)}\hat{p}_T{}^{(2)}) + \tfrac{1}{2}n_4{}^*],$$

$$\hat{p}_t{}^{(2)} = \frac{1}{N}[n_1{}^*\hat{p}_t{}^{(2)} + n_3{}^*(2\hat{p}_T{}^{(1)}\hat{p}_T{}^{(2)}\hat{p}_t{}^{(2)} + 2\hat{p}_T{}^{(1)}\hat{p}_t{}^{(2)}\hat{p}_t{}^{(2)}) \div (4\hat{p}_{t'}^{(1)}\hat{p}_T{}^{(2)}\hat{p}_t{}^{(2)} + 2\hat{p}_T{}^{(1)}\hat{p}_t{}^{(2)}\hat{p}_t{}^{(2)} + \hat{p}_{t'}^{(1)}\hat{p}_T{}^{(2)}\hat{p}_T{}^{(2)}) + \tfrac{1}{2}n_4{}^* + n_5{}^*].$$

If these equations are solved by numerical methods, one would normally only use two of these equations, together with the observation that

$$\hat{p}_T{}^{(1)} + \hat{p}_{t'}^{(1)} = \hat{p}_T{}^{(2)} + \hat{p}_t{}^{(2)} = 1.$$

The functional iteration procedure or the Newton–Raphson procedure examined in Haberman (1971) may be used to find the maximum likelihood estimates.

4.3. *The case of one pair of alleles when the data are incomplete.* As in Section 4.1, suppose that the phenotype is determined by a single pair of alleles, that random mating is present, and that the distribution of alleles is the same for both parents. Also assume that for each $i \in A$, $p_i > 0$. Suppose, however, that some phenotypes cannot be observed or can only rarely be observed. This situation can arise if some phenotype is lethal. If J is the set of observable phenotypes and

$$I = \bigcup_{j \in J} J_j,$$

then the probability that an individual in the sample has allele $i \in A$ from the father and allele $i' \in A$ from the mother is

$$\pi_{ii'} = p_i p_{i'} / \sum_{(d,d') \in I} p_d p_{d'}, \tag{4.13}$$

provided $(i, i') \in I$. The hypothesis that (4.13) is satisfied is equivalent to the

hypothesis that for some α and $\{\beta_i : i \in A\}$,

$$\mu_{ii'} = \alpha + \beta_i + \beta_{i'}, \qquad (i, i') \in I, \tag{4.14}$$

and

$$\sum_{i \in A} \beta_i = 0. \tag{4.15}$$

The set $\mathscr{M}$ of $\boldsymbol{\mu}$ which satisfy (4.14) and (4.15) is a linear manifold spanned by the vectors $\boldsymbol{\mu}^{(t)}$, $t \in A$, where

$$\mu_{ii'}^{(t)} = \delta_{it} + \delta_{i't} \qquad \text{for} \quad (i, i') \in I.$$

If $K_t = \{i : (t, i) \in I\}$, then the maximum likelihood equations become

$$\begin{aligned} \sum_{i \in K_t} m_{ti}(\hat{\boldsymbol{\mu}}) &= N\hat{p}_t(\textstyle\sum_{i \in K_t} \hat{p}_i)/\sum_{(i,i') \in I} \hat{p}_i\hat{p}_{i'} \\ &= \textstyle\sum_{i \in K_t} m_{ti}(\boldsymbol{\mu} \mid \mathbf{n}^*) \\ &= \textstyle\sum_{j \in J} n_j^*(\sum_{(t,i) \in J_j} \hat{p}_t\hat{p}_i)/(\sum_{(i,i') \in J_j} \hat{p}_i\hat{p}_{i'}), \qquad t \in A. \end{aligned}$$

Provided no K_t is empty,

$$\hat{p}_t = \frac{\sum_{j \in J} n_j^*(\sum_{(t,i') \in J_j} \hat{p}_t\hat{p}_{i'})/(\sum_{(i,i') \in J_j} \hat{p}_i\hat{p}_{i'})}{N(\sum_{i \in K_t} \hat{p}_i)/\sum_{(i,i') \in I} \hat{p}_i\hat{p}_{i'}}, \qquad t \in A.$$

To illustrate use of this formula, suppose that a random sample is taken of N subjects belonging to the O, A, and B blood groups. As in Section 1, $J_O = \{OO\}$, $J_A = \{AO, OA, AA\}$, and $J_B = \{BO, OB, BB\}$; however, J is now $\{O, A, B\}$ and $I = \{O, A, B\} \times \{O, A, B\} - \{AB, BA\}$. Since $K_O = \{O, A, B\}$, $K_A = \{O, A\}$, and $K_B = \{O, B\}$,

$$p_O = \frac{n_O^* + n_A^*\hat{p}_O/(\hat{p}_A + 2\hat{p}_O) + n_B\hat{p}_O/(\hat{p}_B + 2\hat{p}_O)}{N/(1 - 2\hat{p}_A\hat{p}_B)},$$

$$p_A = \frac{n_A^*/(\hat{p}_A + 2\hat{p}_O)}{N/(1 - 2\hat{p}_A\hat{p}_B)},$$

$$p_B = \frac{n_B^*/(\hat{p}_B + 2\hat{p}_O)}{N/(1 - 2\hat{p}_A\hat{p}_B)}.$$

5. Multiple solutions of the maximum likelihood equations. Although any maximum likelihood estimate $\hat{\boldsymbol{\mu}}$ of the vector $\boldsymbol{\mu}$ must satisfy (2.3), it does not necessarily follow that (2.3) has a solution or that a given solution of (2.3) is the maximum likelihood estimate $\hat{\boldsymbol{\mu}}$. For example, the likelihood equation for the *ABO* blood group can only lead to a maximum likelihood estimate $\hat{\boldsymbol{\mu}} \in \mathscr{M}$ if (4.5), (4.6), and (4.7) are satisfied for $\hat{p}_O$, $\hat{p}_A$, and $\hat{p}_B$ positive. This cannot occur if $n_{AB}^* = n_B^* = 0$ and $n_O^* > 0$ or $n_A^* > 0$ since (4.7) then implies that $\hat{p}_B = 0$. Even if the restriction that $\hat{p}_O$, $\hat{p}_A$, and $\hat{p}_B$ be positive were removed, all problems would not be solved, for if $n_{AB}^* = n_O^* = 0$ and $n_A^* = n_B^* = 1$, then the maximum likelihood equations would be satisfied by $(\hat{p}_O, \hat{p}_A, \hat{p}_B)$ equal to either $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ or $(0, \frac{1}{2}, \frac{1}{2})$.

As another example, suppose that for some genetic trait involving one pair of alleles, $A = \{1, 2\}$ and $J = \{a, b\}$. Suppose $J_a = \{(1, 2), (2, 1)\}$, $J_b = \{(1, 1), (2, 2)\}$,

$n_a^*/N = \frac{4}{9}$, and $n_b^*/N = \frac{5}{9}$. Then the maximum likelihood equations are

$$\hat{p}_1 = \tfrac{2}{9} + \frac{5\hat{p}_1^2}{9(\hat{p}_1^2 + \hat{p}_2^2)}$$

and

$$\hat{p}_2 = \tfrac{2}{9} + \frac{5\hat{p}_2^2}{9(\hat{p}_1^2 + \hat{p}_2^2)}.$$

These equations are satisfied if $(\hat{p}_1, \hat{p}_2)$ is $(\frac{1}{3}, \frac{2}{3})$, $(\frac{1}{2}, \frac{1}{2})$, or $(\frac{2}{3}, \frac{1}{3})$. The log likelihood for $(\hat{p}_1, \hat{p}_2) = (\frac{1}{2}, \frac{1}{2})$ is

$$N\{\tfrac{4}{9}\log[2(\tfrac{1}{2})(\tfrac{1}{2})] + \tfrac{5}{9}\log[(\tfrac{1}{2})^2 + (\tfrac{1}{2})^2]\} = -N\log 2,$$

while the log likelihood for either $(\hat{p}_1, \hat{p}_2) = (\frac{1}{3}, \frac{2}{3})$ or $(\hat{p}_1, \hat{p}_2) = (\frac{2}{3}, \frac{1}{3})$ is

$$N\{\tfrac{4}{9}\log[2(\tfrac{1}{3})(\tfrac{2}{3})] + \tfrac{5}{9}\log[(\tfrac{1}{3})^2 + (\tfrac{2}{3})^2]\} = N[\tfrac{4}{9}\log(\tfrac{4}{9}) + \tfrac{5}{9}\log(\tfrac{5}{9})],$$

which is greater than $-N\log 2$.

These examples illustrate two problems which may appear when maximum likelihood estimates are to be determined. The first problem is that some solutions of (2.3) do not maximize the log likelihood. This difficulty is related to the fact that the log likelihood need not be concave. The second problem is that (2.7) need have no solution. As shown in the next section, under rather general conditions, this problem does not arise if all observed frequencies n_j^*, $j \in J$, are positive.

To investigate the possibility that (2.3) may have more than one solution, it is necessary to examine the second differential of the log-likelihood function $l(\mathbf{n}^*, \boldsymbol{\mu})$ for Poisson observations. This differential is a linear function on $\mathscr{M}$ such that

$$dl_{\mu+\eta}(\mathbf{n}^*, \boldsymbol{\nu}) = dl_{\mu}(\mathbf{n}^*, \boldsymbol{\nu}) + d^2l_{\mu}(\mathbf{n}^*, \boldsymbol{\nu}, \boldsymbol{\eta}) + \mathscr{O}(\boldsymbol{\nu}, \boldsymbol{\eta}), \tag{5.1}$$

$$\boldsymbol{\nu} \in \mathscr{M},\ \boldsymbol{\eta} \in \mathscr{M},$$

where

$$\sup_{\|\boldsymbol{\nu}\|=1} |\mathscr{O}(\boldsymbol{\nu}, \boldsymbol{\eta})| \to 0$$

as $\boldsymbol{\eta} \to 0$. If one proceeds as with (2.7), one finds that

$$\begin{aligned} d^2l_{\mu}(\mathbf{n}^*, \boldsymbol{\nu}, \boldsymbol{\eta}) = {} & \textstyle\sum_{j\in J} n_j^*(\sum_{i\in J_j} e^{\mu_i}\nu_i\eta_i)/(\sum_{i\in J_j} e^{\mu_i}) \\ & - \textstyle\sum_{j\in J} n_j^*(\sum_{i\in J_j} e^{\mu_i}\nu_i)(\sum_{i\in J_j} e^{\mu_i}\eta_i)/(\sum_{i\in J_j} e^{\mu_i})^2 \\ & - \textstyle\sum_{i\in I} e^{\mu_i}\nu_i\eta_i, \qquad \boldsymbol{\nu} \in \mathscr{M},\ \boldsymbol{\eta} \in \mathscr{M}. \end{aligned} \tag{5.2}$$

Since the conditional covariance of $(\mathbf{n}, \boldsymbol{\nu})$ and $(\mathbf{n}, \boldsymbol{\eta})$ given $\mathbf{n}^*$ is

$$\begin{aligned} \mathrm{Cov}_{\mu}(\boldsymbol{\nu}, \boldsymbol{\eta} \mid \mathbf{n}^*) = {} & \textstyle\sum_{j\in J} n_j^*[(\sum_{i\in J_j} e^{\mu_i}\nu_i\eta_i)/(\sum_{i\in J_j} e^{\mu_i}) \\ & - (\textstyle\sum_{i\in J_j} e^{\mu_i}\nu_i)(\sum_{i\in J_j} e^{\mu_i}\eta_i)/(\sum_{i\in J_j} e^{\mu_i})^2] \end{aligned} \tag{5.3}$$

and the unconditional covariance of $(\mathbf{n}, \boldsymbol{\nu})$ and $(\mathbf{n}, \boldsymbol{\eta})$ for independent Poisson observations is

$$\mathrm{Cov}_{\mu}(\boldsymbol{\nu}, \boldsymbol{\eta}) = \textstyle\sum_{i\in I} e^{\mu_i}\nu_i\eta_i, \tag{5.4}$$

(5.2) may be written as

$$d^2 l_\mu(\mathbf{n}^*, \boldsymbol{\nu}, \boldsymbol{\eta}) = \mathrm{Cov}_\mu(\boldsymbol{\nu}, \boldsymbol{\eta} \mid \mathbf{n}^*) - \mathrm{Cov}_\mu(\boldsymbol{\nu}, \boldsymbol{\eta}), \qquad \boldsymbol{\nu} \in \mathscr{M}, \boldsymbol{\eta} \in \mathscr{M}. \tag{5.5}$$

In (5.5), $\mathrm{Cov}_\mu(\boldsymbol{\nu}, \boldsymbol{\eta})$ is a positive definite quadratic form and $\mathrm{Cov}_\mu(\boldsymbol{\nu}, \boldsymbol{\eta} \mid \mathbf{n}^*)$ is a nonnegative definite quadratic form. The difference between these forms need not be negative definite. Consequently, $l(\mathbf{n}^*, \boldsymbol{\mu})$ need not be concave. Thus the log likelihood may have more than one critical point and (2.3) may have more than one solution, although any solution $\hat{\boldsymbol{\mu}}$ of (2.3) such that $d^2 l_{\hat{\mu}}(\mathbf{n}^*, \boldsymbol{\nu}, \boldsymbol{\eta})$ is negative definite must be at least a relative maximum of $l(\mathbf{n}^*, \boldsymbol{\mu})$.

The possibility that (2.3) may have multiple solutions may also be explored by examination of the second differential $d^2 l_\mu^{(m)}(\mathbf{n}^*, \boldsymbol{\nu}, \boldsymbol{\eta})$ for $\boldsymbol{\nu} \in \mathscr{M} - \mathscr{N}$ and $\boldsymbol{\eta} \in \mathscr{M} - \mathscr{N}$. In this case, one finds that

$$d^2 l_{\mu'}^{(m)}(\mathbf{n}^*, \boldsymbol{\nu}, \boldsymbol{\eta}) = \mathrm{Cov}_{\mu'}^{(m)}(\boldsymbol{\nu}, \boldsymbol{\eta} \mid \mathbf{n}^*) - \mathrm{Cov}_{\mu'}^{(m)}(\boldsymbol{\nu}, \boldsymbol{\eta}), \tag{5.6}$$
$$\boldsymbol{\nu} \in \mathscr{M} - \mathscr{N}, \boldsymbol{\eta} \in \mathscr{M} - \mathscr{N},$$

where $\mathrm{Cov}_{\mu'}^{(m)}(\boldsymbol{\nu}, \boldsymbol{\eta} \mid \mathbf{n}^*) = \mathrm{Cov}_\mu(\boldsymbol{\nu}, \boldsymbol{\eta} \mid \mathbf{n}^*)$ if $\boldsymbol{\mu}$ is defined by (3.2) and

$$\mathrm{Cov}_{\mu'}^{(m)}(\boldsymbol{\nu}, \boldsymbol{\eta}) = \sum_{i \in I} e^{\mu_i} \nu_i \eta_i - \sum_{k=1}^{\tau} \left(\sum_{i \in I_k} e^{\mu_i} \nu_i\right)\left(\sum_{i \in I_k} e^{\mu_i} \eta_i\right)/N_k$$

is the covariance of $(\mathbf{n}, \boldsymbol{\nu})$ and $(\mathbf{n}, \boldsymbol{\eta})$ under multinomial sampling (see Haberman (1970 and 1974)). Once again, the second differential is the difference between a nonnegative definite and a positive definite quadratic form and $l^{(m)}(\mathbf{n}^*, \boldsymbol{\mu}')$ need not be concave. If $\hat{\boldsymbol{\mu}}$ satisfies (2.3) and $\boldsymbol{\mu}' = P_{\mathscr{M}-\mathscr{N}} \hat{\boldsymbol{\mu}}$ is such that $d^2 l_{\hat{\mu}'}^{(m)}(\mathbf{n}^*, \boldsymbol{\nu}, \boldsymbol{\eta})$ is negative definite, then $\hat{\boldsymbol{\mu}}'$ is at least a relative maximum of $l^{(m)}(\mathbf{n}^*, \boldsymbol{\mu}')$.

The problem of multiple solutions of the maximum likelihood equations disappears in the perfect observation case since $\mathrm{Cov}_\mu(\boldsymbol{\nu}, \boldsymbol{\eta} \mid \mathbf{n}^*)$ is identically 0 when $\mathbf{n} = \mathbf{n}^*$. In this case, if (2.3) has a solution $\hat{\boldsymbol{\mu}}$, this solution is the unique maximum likelihood estimate of $\boldsymbol{\mu}$ for both multinomial and Poisson samples. In addition, both Fienberg and Larntz (1971) and Chen (1972) present examples where $J \neq I$ but the maximum likelihood estimate $\hat{\boldsymbol{\mu}}$ is unique when it exists and is the only solution of the maximum likelihood equations. It will be shown in a later paper that the problem of multiple solutions of (2.3) normally does not lead to serious computational difficulties in cases where moderately large samples are present.

6. Sufficient conditions for existence of a maximum likelihood estimate. If a Fisher-consistent estimate of $\boldsymbol{\mu}$ is available, then one can at least ensure that the log likelihood does have a maximum for some $\hat{\boldsymbol{\mu}} \in \mathscr{M}$ which satisfies (2.7) whenever $n_j^* > 0$ for $j \in J$. In this assertion, we modify Rao's (1965) definition of Fisher-consistency so that an estimate $T(\mathbf{n}^*)$ of $\boldsymbol{\mu}$ is Fisher-consistent if T is a function with range $\mathscr{M}$ such that

$$T(\{\textstyle\sum_{i \in J_j} e^{\mu_i} : j \in J\}) = \boldsymbol{\mu} \tag{6.1}$$

for all $\boldsymbol{\mu} \in \mathscr{M}$ and $T(\mathbf{x})$ is continuous for $\mathbf{x} \in R^J$ such that $x_j > 0$ for $j \in J$. We prove the following theorem:

THEOREM 3. *If there exists a Fisher-consistent estimate of* μ *and if* $n_j^* > 0$ *for all* $j \in J$, *then for some* $\hat{\mu} \in \mathcal{M}$,

$$l(\mathbf{n}^*, \hat{\mu}) = \sup_{\mu \in \mathcal{M}} l(\mathbf{n}^*, \mu)$$

and

$$l^{(m)}(\mathbf{n}^*, P_{\mathcal{M}-\mathcal{N}}\hat{\mu}) = \sup_{\mu' \in \mathcal{M}-\mathcal{N}} l^{(m)}(\mathbf{n}^*, \mu') .$$

PROOF. Observe that when $n_j^* > 0$ for $j \in J$, the set A of $\mathbf{x} \in R^J$ such that

$$\sum_{j \in J} (n_j^* x_j - e^{x_j} - \log n_j^*!) \geqq l(\mathbf{n}^*, \mathbf{0})$$

is closed, nonempty, and bounded. The claim follows since the two sides are equal when

$$x_j = \log \sum_{i \in J_j} e^0 , \qquad j \in J ,$$

and the summand

$$n_j^* x_j - e^{x_j} - \log n_j^*!$$

is bounded above for all x_j and approaches $-\infty$ as $|x_j| \to \infty$ (see Birch (1963)). Since T is continuous, the set

$$B = \{T(\{e^{x_j} : j \in J\}) : \mathbf{x} \in A\}$$

is also closed, nonempty, and bounded.

If $\mu \in \mathcal{M}$ and $l(\mathbf{n}^*, \mu) \geqq l(\mathbf{n}^*, \mathbf{0})$, then $\mathbf{x} \in A$, where $x_j = \log (\sum_{i \in J_j} e^{\mu_i})$. By (6.1), $\mu \in B$. Since $B \subset \mathcal{M}$, B is thus the set of $\mu \in \mathcal{M}$ such that $l(\mathbf{n}^*, \mu) \geqq l(\mathbf{n}^*, \mathbf{0})$. Since the set is closed and bounded, the supremum of $l(\mathbf{n}^*, \mu)$ must be achieved at a point $\hat{\mu} \in B$ which satisfies (2.3). As observed in Section 3, the supremum of the log likelihood $l^{(m)}(\mathbf{n}^*, \mu')$ for multinomial sampling must be achieved at $P_{\mathcal{M}-\mathcal{N}}\hat{\mu}$. □

This result concerning the existence of a maximum likelihood estimate applies to the blood group example of Sections 1 and 4.1 since $\{\log N + \log \bar{p}_i + \log \bar{p}_{i'} : (i, i') \in I\}$ is Fisher-consistent, where

$$N = n_O^* + n_A^* + n_B^* + n_{AB}^* , \tag{6.2}$$

$$\bar{p}_O = (n_O^*/N)^{\frac{1}{2}} , \tag{6.3}$$

$$\bar{p}_A = ((n_O^* + n_A^*)/N)^{\frac{1}{2}} - \bar{p}_O , \tag{6.4}$$

and

$$\bar{p}_B = ((n_O^* + n_B^*)/N)^{\frac{1}{2}} - \bar{p}_O . \tag{6.5}$$

The estimates $\bar{p}_O$, $\bar{p}_A$, and $\bar{p}_B$ are the Wiener estimates of gene frequencies given in Elandt–Johnson (1971, page 397). Thus a maximum likelihood estimate of (p_O, p_A, p_B) exists such that $0 < p_i < 1$ for $i \in \{O, A, B\}$ if n_O^*, n_A^*, n_B^*, and n_{AB}^* are all positive.

7. Conclusion. This paper has considered construction of maximum likelihood equations for a specific class of models. These equations have been shown to be the same under all sampling schemes considered, and maximum likelihood equations have been related to conditional and unconditional expected values.

Some properties of solutions of these equations have been explored, but numerical methods of computation of maximum likelihood estimates and asymptotic properties of these estimates have not been considered. These topics will be the subjects of future papers. In some special cases, they are examined by Chen (1972) and Haberman (1971).

REFERENCES

[1] BIRCH, M. W. (1963). Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc. Ser. B* **25** 220–233.

[2] CEPPELLINI, R., SINISCALCO, S. and SMITH, C. A. B. (1955). The estimation of gene frequencies in a random-mating population. *Ann. Human Genetics* **20** 97–115.

[3] CHEN, T. (1972). Mixed-up frequencies in contingency tables. Ph. D. dissertation, Univ. of Chicago.

[4] ELANDT-JOHNSON, R. C. (1971). *Probability Models and Statistical Methods in Genetics.* Wiley, New York.

[5] FIENBERG, S. E. and LARNTZ, K. L. (1971). Some models for individual-group comparisons and group behavior. *Psychometrika* **36** 349–367.

[6] HABERMAN, S. J. (1970). The general log-linear model. Ph. D. dissertation, Univ. of Chicago.

[7] HABERMAN, S. J. (1971). Tables based on imperfect observation. Invited paper, 1971 ENAR meeting, Pennsylvania State Univ.

[8] HABERMAN, S. J. (1973). Log-linear models for frequency data: sufficient statistics and likelihood equations. *Ann. Statist.* **1** 617–632.

[9] HABERMAN, S. J. (1974). *The Analysis of Frequency Data.* To be published by University of Chicago Press.

[10] KOLER, R. D., JONES, R. T., WASI, P. and POOTRUKUL, S. (1971). Genetics of haemoglobin *H* and α-thalassemia. *Ann. Human Genetics* **34** 371–377.

[11] RAO, C. R. (1965). *Linear Statistical Inference and Its Applications.* Wiley, New York.

[12] YASUDA, N. and KIMURA, M. (1968). A gene counting method of maximum likelihood in *ABO* and *ABO*-like systems. *Ann. Human Genetics* **31** 409–420.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
1118 EAST 58TH STREET
CHICAGO, ILLINOIS 60637