The Annals of Statistics
1977, Vol. 5, No. 6, 1124–1147

PRODUCT MODELS FOR FREQUENCY TABLES INVOLVING INDIRECT OBSERVATION[1]

BY SHELBY J. HABERMAN

University of Chicago

Frequency tables are often encountered which cannot be directly observed. Examples occur in gene frequency estimation problems, latent structure analysis, epidemiology, and studies of group interactions. A class of models is proposed for these various applications. Maximum likelihood equations are derived, together with methods for their solution. Large-sample properties of these estimates are studied.

1. Introduction. In many probability models for a frequency table $\mathbf{n} = \{n_i : i \in I\}$, it is assumed that $\mathbf{n}$ consists of $K \geqq 1$ independent multinomial vectors $\{n_i : i \in I_k\}$, $1 \leqq k \leqq K$, with respective sample sizes $N_k > 0$ and associated probabilities $\{p_i(\boldsymbol{\pi}) : i \in I_k\}$ such that

$$p_i(\boldsymbol{\pi}) = d_i \prod_{h \in H} \pi_h^{c(h,i)}, \qquad i \in I. \tag{1.1}$$

Here H is a nonempty finite index set, $d_i > 0$, the $c(h, i)$ are known nonnegative integers, and the π_h are unknown nonnegative parameters such that $\boldsymbol{\pi} = \{\pi_h : h \in H\}$ is in some affine subspace Θ of the space R^H of functions from H to R. To assure the existence of maximum likelihood estimates, it is assumed that the set $\Theta_+ = \{\boldsymbol{\pi} \in \Theta : \pi_h \geqq 0 \ \forall\, h \in H\}$ is nonempty and compact. To ensure that Θ_+ corresponds to a set of probabilities $p_i(\boldsymbol{\pi})$, $i \in I$, it is assumed that if $\boldsymbol{\pi} \in \Theta_+$, then

$$\sum_{i \in I_k} p_i(\boldsymbol{\pi}) = 1, \qquad 1 \leqq k \leqq K.$$

If $\mathbf{n}$ can be observed, then estimation of $\boldsymbol{\pi}$ is a relatively straightforward procedure. However, estimation is somewhat more difficult if instead of $\mathbf{n}$, only $\mathbf{n}^* = \{n_j^* : j \in J\}$ is observed, where

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

the J_j are disjoint nonempty sets with union I, and each J_j is contained in some I_k.

To illustrate the type of problems in which (1.1) holds and $\mathbf{n}$ cannot be observed, the following four examples from genetics and latent-structure analysis may be considered.

EXAMPLE 1. *A general model for estimation of gene frequencies.* Suppose that a random sample of size N is taken from a population in Hardy–Weinberg

Received January 1974; revised January 1977.

[1] Support for this research has been provided in part by Research Grant No. NSF GP 32037X from the Division of Mathematical, Physical, and Engineering Sciences, and NSF GS 31967X from the Division of Social Sciences, National Science Foundation.

AMS 1970 *subject classifications.* Primary 62F10; Secondary 62P10, 62P25.

Key words and phrases. Nonlinear programming, indirect observation, contingency table, genetic model, latent-structure analysis, maximum likelihood estimation.

equilibrium (see Elandt-Johnson, 1971, pages 54–82). Suppose that pairs of genes at loci $g \in G$ determine the phenotype $j \in J$ of a population member. Suppose each gene at locus g has possible alleles H_g, and let the gene frequency of allele $h \in H_g$ be π_h. Provided that the H_g are disjoint, the π_h are nonnegative unknown quantities which belong to the affine space

$$\Theta = \{\boldsymbol{\pi} \in R^H : \textstyle\sum_{h \in H_g} \pi_h = 1\} . \tag{1.3}$$

A member of the population has genotype $\mathbf{i} = \{\langle i(1, g), i(2, g)\rangle : g \in G\}$, where $i(1, g)$ is the gene at locus g received from the male parent and $i(2, g)$ is the corresponding gene received from the female parent. If the effects of linkage can be ignored, then the probability is

$$p_{\mathbf{i}}(\boldsymbol{\pi}) = \textstyle\prod_{g \in G} \pi_{i(1,g)} \pi_{i(2,g)} \tag{1.4}$$

that a randomly selected population member has genotype i. It should be noted that the distinction between parents in the definition of genotype is not normally made by geneticists. It is adopted here for simplicity of presentation.

If $n_{\mathbf{i}}$ denotes the number of members of the sample with genotype $\mathbf{i}$ and if

$$I = \textstyle\prod_{g \in G} (H_g \times H_g) ,$$

then $\mathbf{n} = \{n_{\mathbf{i}} : i \in I\}$ has a multinomial distribution with sample size N and probabilities $\mathbf{p}(\boldsymbol{\pi}) = \{p_i(\boldsymbol{\pi}) : i \in I\}$.

Estimation of the probabilities π_h would not be complicated if $\mathbf{n}$ were observable; however, in practice only phenotypes can be observed. Suppose that phenotype j is obtained if the genotype $\mathbf{i}$ is in J_j, where $\mathbf{i}$ and $\mathbf{d}$ are both in J_j if for each g, $i(1, g) = d(1, g)$ and $i(2, g) = d(2, g)$ or $i(1, g) = d(2, g)$ and $i(2, g) = d(1, g)$. The n_j^*, the number of subjects in the sample with phenotype j, satisfies (1.2).

EXAMPLE 2. *A two-loci model.* Koler, Jones, Wasi and Pootrukul (1971) proposed a two-loci model for decreased synthesis of human hemoglobin α-chains. In this model, the first locus has alleles T' and t', while the second locus has alleles T and t. Five phenotypes are observed. The phenotypes, together with corresponding genotypes, are given in Table 1. Thus $G = \{1, 2\}$, $H = \{T', t', T, t\}$, the J_j are defined by means of Table 1,

$$p_{\mathbf{i}}(\boldsymbol{\pi}) = \pi_{i(1,1)}\pi_{i(1,2)}\pi_{i(2,1)}\pi_{i(2,2)} ,$$

TABLE 1
Phenotypes and genotypes for the two-locus model

Phenotype symbol	Phenotype name	Corresponding genotypes
1	Normal	$\{\langle T', T'\rangle, \langle x, y\rangle\}$, $x, y = T$ or t
2	Silent carrier	$\{\langle T', t'\rangle, \langle T, T\rangle\}$, $\{\langle t', T'\rangle, \langle T, T\rangle\}$
3	α-thalassemia trait	$\{\langle T', t'\rangle, \langle x, y\rangle\}$, $\{\langle t', T'\rangle, \langle x, y\rangle\}$, x or $y \neq T$, or $\{\langle t', t'\rangle, \langle T, T\rangle\}$
4	*Hb H* disease	$\{\langle t', t'\rangle, \langle T, t\rangle\}$, $\{\langle t', t'\rangle, \langle t, T\rangle\}$
5	Hydrops foetalis	$\{\langle t', t'\rangle, \langle t, t\rangle\}$

and Θ consists of $\boldsymbol{\pi} \in R^H$ such that $\pi_{T'} + \pi_{t'} = \pi_T + \pi_t = 1$. The observed table is $\mathbf{n}^* = \{n_j^* : 1 \leqq j \leqq 5\}$, and $I = (\{T', t'\} \times \{T', t'\}) \times (\{T, t\} \times \{T, t\})$.

EXAMPLE 3. *Latent-structure analysis.* Members of a population have observed (manifest) characteristics $A(u) \in L_u$, $u \in U$, and unobserved (latent) charateristics $X(t) \in K_t$, $t \in T$. Thus if a random sample of size N is taken from the population, one may observe the number $n_{\mathbf{j}}^*$ of subjects with $A(u) = j(u) \in L_u$, $u \in U$; however, the number $n_{\mathbf{kj}}$ of subjects with $X(t) = k(t) \in K_t$, $t \in T$, and $A(u) = j(u) \in L_u$, $u \in U$, cannot be observed. In this example, (1.2) holds if

$$I = (\textstyle\prod_{t \in T} K_t) \times (\prod_{u \in U} L_u)\,,$$
$$J = \textstyle\prod_{u \in U} L_u\,,$$

and

$$J_{\mathbf{j}} = \{\langle \mathbf{k}, \mathbf{j} \rangle : \mathbf{k} \in \textstyle\prod_{t \in T} K_t\}\,, \qquad \mathbf{j} \in J\,.$$

The classical local independence assumption is made that for a randomly chosen population member, the $A(u)$, $u \in U$, are conditionally independent given $\mathbf{X} = \{X(t) : t \in T\}$ (see Lazarsfeld and Henry (1968)). Thus if $\pi_{\mathbf{k}} = P\{\mathbf{X} = \mathbf{k}\}$ and $\pi_{\mathbf{k}ug} = P\{A(u) = g \mid \mathbf{X} = \mathbf{k}\}$, then $\mathbf{n}$ has a multinomial distribution with sample size N and probabilities $p(\boldsymbol{\pi}) = \{p_i(\boldsymbol{\pi}) : i \in I\}$ such that if $\mathbf{i} = \langle \mathbf{k}, \mathbf{j} \rangle$, then

$$p_{\mathbf{i}}(\boldsymbol{\pi}) = \pi_{\mathbf{k}} \textstyle\prod_{u \in U} \pi_{\mathbf{k}uj(u)}\,. \tag{1.5}$$

The set H may be defined as $K \cup \{\langle \mathbf{k}, u, g \rangle : \mathbf{k} \in K, g \in L_u, u \in U\}$, where $K = \prod_{t \in T} K_t$. The vector $\boldsymbol{\pi}$ is then in the space Θ' of $\boldsymbol{\pi} \in R^H$ such that

$$\textstyle\sum_{k \in K} \pi_{\mathbf{k}} = 1 \quad \text{and} \quad \sum_{g \in L_u} \pi_{\mathbf{k}ug} = 1\,, \qquad \mathbf{k} \in K,\ u \in U\,. \tag{1.6}$$

In traditional latent-structure analysis, no further assumptions are made concerning $\boldsymbol{\pi}$; however, Goodman (1974) considers possible added linear restrictions on $\boldsymbol{\pi}$. For instance, for subsets N and B_c, $c \in C$, of H, it may be assumed that $\boldsymbol{\pi} \in \Theta$, where $\boldsymbol{\pi} \in \Theta$ if $\boldsymbol{\pi} \in \Theta'$, $\pi_h = 0$ for $h \in N$, and $\pi_h = \pi_{h'}$ if for some c, $c \in C$, h and h' are in B_c.

EXAMPLE 4. *An example with two latent variables.* Goodman (1974) gives considerable attention to a model with two dichotomous latent variables and four dichotomous manifest variables. In the language of Example 3, $T = \{1, 2\}$, $U = \{1, 2, 3, 4\}$, and $K_1 = K_2 = L_1 = L_2 = L_3 = L_4 = \{1, 2\}$. In addition to the local independence assumption, it is aśsumed that $P\{A(1) = j(1) \mid X(1) = k(1), X(2) = k(2)\}$ and $P\{A(3) = j(3) \mid X(1) = k(1), X(2) = k(2)\}$ are independent of $k(2)$, and it is assumed that $P\{A(2) = j(2) \mid X(1) = k(1), X(2) = k(2)\}$ and $P\{A(4) = j(4) \mid X(1) = k(1), X(2) = k(2)\}$ are independent of $k(1)$. Thus $\boldsymbol{\pi} \in \Theta$, where Θ consists of $\boldsymbol{\pi} \in \Theta'$ such that $\pi_{\mathbf{k}1j(1)} = \pi_{\mathbf{k}'1j(1)}$ if $k(1) = k'(1)$, $\pi_{\mathbf{k}3j(3)} = \pi_{\mathbf{k}'3j(3)}$ if $k(1) = k'(1)$, $\pi_{\mathbf{k}2j(2)} = \pi_{\mathbf{k}'2j(2)}$ if $k(2) = k'(2)$, and $\pi_{\mathbf{k}4j(4)} = \pi_{\mathbf{k}'tj(4)}$ if $k(2) \doteq k'(2)$.

Maximum likelihood methods provide an effective means for estimation of $\boldsymbol{\pi}$ and $\mathbf{p}(\boldsymbol{\pi}) = \{p_i(\boldsymbol{\pi}) : i \in I\}$. As shown in Section 2, the maximum likelihood estimate $\hat{\boldsymbol{\pi}}$ of $\boldsymbol{\pi}$ may be defined in terms of conditional expected values $m_h(\boldsymbol{\pi} \mid \mathbf{n}^*)$ of the random variable

$$y_h = \sum_{i \in I} c(h, i) n_i \tag{1.7}$$

given the observation $\mathbf{n}^*$. Let

$$p_j^*(\boldsymbol{\pi}) = \sum_{i \in J_j} p_i(\boldsymbol{\pi}) . \tag{1.8}$$

Then

$$m_h(\boldsymbol{\pi} \mid \mathbf{n}^*) = \sum_{j \in J} n_j^* [\sum_{i \in J_j} c(h, i) p_i(\boldsymbol{\pi}) / p_j^*(\boldsymbol{\pi})] , \tag{1.9}$$

where the convention $0/0 = 0$ is used. Let $\Omega = \{\mathbf{z} - \mathbf{w} : \mathbf{z}, \mathbf{w} \in \Theta\}$, and let $\Omega^\perp$ be the orthogonal complement of Ω.

Let $(\partial/\partial\pi_h) m_h(\hat{\boldsymbol{\pi}} \mid n^*)$ be the value at $\hat{\boldsymbol{\pi}}$ of the partial derivative of $m_h(\cdot \mid \mathbf{n}^*)$ with respect to π_h. Given these definitions, some $\mathbf{b} \in \Omega^\perp$ exists such that

$$m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) = b_h \hat{\pi}_h , \qquad h \in H , \tag{1.10}$$

and

$$b_h \geqq \frac{\partial}{\partial \pi_h} m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) , \qquad \hat{\boldsymbol{\pi}}_h = 0 , \qquad h \in H . \tag{1.11}$$

In the case of the genetic models of Examples 1 and 2, (1.10) is shown to reduce to the gene counting equations

$$\hat{\boldsymbol{\pi}}_h = \frac{1}{2N} m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) , \qquad h \in H , \tag{1.12}$$

of Ceppellini, Siniscalco and Smith (1955). In the case of the latent structure models of Examples 3 and 4, the resulting equations depend somewhat on the specific choice of Θ. The examples in Section 2 provide a general procedure for obtaining equations such as those in Goodman (1974).

Equation (1.10) is not necessarily sufficient to determine $\hat{\boldsymbol{\pi}}$. More than one maximum likelihood estimate of $\boldsymbol{\pi}$ may exist or a solution of (1.10) may fail to maximize the likelihood function. Consequently, conditions are provided in Section 2 to assist in identification of solutions of (1.10) which are at least relative maxima of the likelihood function. As shown in Section 3, these problems involving multiple maxima are often relatively minor if the sample sizes N_k are large.

In Section 3, asymptotic properties of $\hat{\boldsymbol{\pi}}$ are discussed under the condition that $N \to \infty$ and $N_k/N \to \tau_k$, $1 \leqq k \leqq K$, where

$$N = \sum_{k=1}^K N_k .$$

These properties are shown to depend on the limit $\mathbf{m}(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)$ of the expected value of $N^{-1}\mathbf{m}(\boldsymbol{\pi} \mid \mathbf{n}^*)$ and on the limit $C(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)$ of the expected value of $N^{-1}C(\boldsymbol{\pi} \mid \mathbf{n}^*)$. Here

$$\boldsymbol{\mu}^* = \{\tau_k p_j^*(\boldsymbol{\pi}) : j \in A_k, 1 \leqq k \leqq K\} ,$$
$$A_k = \{j \in J : J_j \subset I_k\} , \quad 1 \leqq k \leqq K ,$$

and $C(\boldsymbol{\pi}\mid\mathbf{n}^*)$ is the linear transformation on R^H with matrix elements

$$(1.13)\quad \begin{aligned}[C(\boldsymbol{\pi}\mid\mathbf{n}^*)]_{hh'} = \textstyle\sum_{j\in J} A_j^*\{&\textstyle\sum_{i\in J_j} c(h,i)c(h',i)p_i(\boldsymbol{\pi})/p_j^*(\boldsymbol{\pi})\\ &- [\textstyle\sum_{i\in J_j} c(h,i)p_i(\boldsymbol{\pi})/p_j^*(\boldsymbol{\pi})]\\ &\times [\textstyle\sum_{i\in J_j} c(h',i)p_i(\boldsymbol{\pi})/p_j(\boldsymbol{\pi})]\}\,, \qquad h, h'\in H\,.\end{aligned}$$

Thus $C(\boldsymbol{\pi}\mid\mathbf{n}^*)$ is the conditional covariance operator of $\mathbf{y}$ given $\mathbf{n}^*$; that is,

$$(1.14)\quad (\mathbf{w}, C(\boldsymbol{\pi}\mid\mathbf{n}^*)\mathbf{z}) = \operatorname{Cov}[(\mathbf{w},\mathbf{y}),(\mathbf{z},\mathbf{y})\mid\mathbf{n}^*]\,, \qquad \mathbf{w},\mathbf{z}\in R^H\,.$$

Let

$$(1.15)\quad B(\boldsymbol{\pi}\mid\boldsymbol{\mu}^*)\mathbf{z} = \{m_h(\boldsymbol{\pi}\mid\boldsymbol{\mu}^*)z_h\}\,, \qquad \mathbf{z}\in R^H\,.$$

$$(1.16)\quad E(\boldsymbol{\pi}\mid\boldsymbol{\mu}^*) = \Pi^-(\boldsymbol{\pi})[B(\boldsymbol{\pi}\mid\boldsymbol{\mu}^*) - C(\boldsymbol{\pi}\mid\boldsymbol{\mu}^*)]\Pi^-(\boldsymbol{\pi})\,,$$

and

$$(1.17)\quad \begin{aligned}[\Pi^-(\boldsymbol{\pi})\mathbf{z}]_h &= z_h/\pi_h\,, \quad \pi_h\neq 0\,,\\ &= 0\,, \qquad \pi_h = 0\,.\end{aligned}$$

As noted in Section 3, for some $\boldsymbol{\beta}\in\Omega^\perp$,

$$(1.18)\quad m_h(\boldsymbol{\pi}\mid\boldsymbol{\mu}^*) = \beta_h\pi_h\,, \qquad h\in H\,,$$

and

$$(1.19)\quad \frac{\partial}{\partial\pi_h} m_h(\boldsymbol{\pi}\mid\boldsymbol{\mu}^*) \leqq \beta_h\,, \qquad \pi_h\in 0,\ h\in H\,.$$

Let $\Omega'(\boldsymbol{\pi})$ consist of all $\mathbf{x}\in\Omega$ such that $x_h = 0$ whenever $\pi_h = 0$. If strict inequality holds in (1.19) whenever $\pi_h = 0$ and if

$$(1.20)\quad (\mathbf{z}, E(\boldsymbol{\pi}\mid\boldsymbol{\mu}^*)\mathbf{z}) > 0\,, \qquad \mathbf{z}\neq\mathbf{0},\ \mathbf{z}\in\Omega'(\boldsymbol{\pi})\,,$$

then there exist functions $\hat{\boldsymbol{\pi}}(\cdot)$ and $\mathbf{b}(\cdot)$ such that (1.10) and (1.11) hold for $\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}}(\mathbf{n}^*)$ and $\mathbf{b} = \mathbf{b}(\mathbf{n}^*)$ and such that $N^{\frac{1}{2}}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi})$ converges to $N(\mathbf{0}, \Sigma(\boldsymbol{\pi}\mid\boldsymbol{\mu}^*))$ in distribution. Here

$$(1.21)\quad \Sigma(\boldsymbol{\pi}\mid\boldsymbol{\mu}^*) = P(\boldsymbol{\pi}\mid\boldsymbol{\mu}^*)[E(\boldsymbol{\pi}\mid\boldsymbol{\mu}^*)]^-[P(\boldsymbol{\pi}\mid\boldsymbol{\mu}^*)]^A\,,$$

$P(\boldsymbol{\pi}\mid\boldsymbol{\mu}^*)$ is the projection on $\Omega'(\boldsymbol{\pi})$ with respect to $E(\boldsymbol{\pi}\mid\boldsymbol{\mu}^*)$, $\{\ \}^-$ denotes any generalized inverse, and $\{\ \}^A$ denotes an adjoint (see Rao and Mitra (1971, pages 3 and 20)). More can be said if $\boldsymbol{\pi}$ is the only element $\mathbf{x}\in\Theta_+$ such that if $\tau_k > 0$ and $j\in A_k$, then $p_j^*(\mathbf{x}) = p_j^*(\boldsymbol{\pi})$. In this case, $\hat{\boldsymbol{\pi}}(\cdot)$ may be defined so that $\hat{\boldsymbol{\pi}}(\mathbf{n}^*)$ maximizes $l(\mathbf{n}^*, \cdot)$. The probability approaches 1 that $\hat{\boldsymbol{\pi}}(\mathbf{n}^*)$ is uniquely defined in this manner.

In other portions of Section 3, alternate regularity conditions are provided for the existence of an asymptotically normal estimate $\hat{\boldsymbol{\pi}}(\mathbf{n}^*)$, estimation of asymptotic covariances is discussed, and tests of goodness of fit are presented.

In Section 4, an unusually stable numerical procedure is presented for computation of solutions of (1.10). It is assumed that a maximum $\bar{\boldsymbol{\pi}}(\mathbf{y})$ of

$$\textstyle\prod_{i\in J}[p_i(\boldsymbol{\pi})]^{y_i}$$

for $\boldsymbol{\pi} \in \Theta_+$ can be readily computed for $\mathbf{y} \in [0, \infty)^I$. Given this assumption, if $\boldsymbol{\pi}^{(0)}$ is an initial estimate of a solution $\hat{\boldsymbol{\pi}}$ of (1.10) and if for $v \geqq 0$,

$$\boldsymbol{\pi}^{(v+1)} = \bar{\boldsymbol{\pi}}(\{n_j{}^* p_i(\boldsymbol{\pi}^{(v)})/p_j{}^*(\boldsymbol{\pi}^{(v)}) : i \in J_j, j \in J\}) ,$$

then every subsequence of $\{\boldsymbol{\pi}^{(v)} : v \geqq 0\}$ contains a subsequence which converges to a $\hat{\boldsymbol{\pi}} \in \Theta_+$ such that (1.10) holds for some $\mathbf{b} \in \Omega^\perp$. If $\boldsymbol{\pi}^{(0)}$ is sufficiently close to an isolated maximum likelihood estimate $\hat{\boldsymbol{\pi}}$, then $\boldsymbol{\pi}^{(v)} \to \hat{\boldsymbol{\pi}}$. This result is especially helpful if the large-sample assumptions of Section 3 hold, for if $\boldsymbol{\pi}^{(0)}$ is a consistent estimate of $\boldsymbol{\pi}$, then $\boldsymbol{\pi}^{(v)}$ converges to the unique maximum likelihood estimate $\hat{\boldsymbol{\pi}}$ with probability approaching 1. The algorithm in this section reduces to the gene counting method of Ceppellini, Siniscalco and Smith (1965) and to the iterative procedures for latent-structure analysis of Goodman (1974).

In Section 5, more rapidly convergent Newton–Raphson and scoring algorithms are considered. They are also shown to be described in terms of conditional and unconditional moments of the y_h. The resulting scoring algorithms correspond to those in Smith (1956) or Anderson (1959).

The results developed in this paper can also be applied to the models for group behavior of Fienberg and Larntz (1971), to the models for mixed-up frequencies in contingency tables considered by Chen (1972) and Chen and Fienberg (1974), and to Cohen's (1971) model for a censored $2 \times 2 \times 2$ table. The Markov models of Anderson and Goodman (1957) are special cases of the models considered in this paper, as are those hierarchical models in Goodman (1970) which have closed-form maximum likelihood estimates. Haberman's (1974, 1976) results concerning indirect observations of frequency tables which obey log-linear models are consistent with results in this paper, but the models in this paper need not correspond to log-linear models and log-linear models need not correspond to models in this paper. Sundberg (1971, 1974) has recently considered indirect observation problems for exponential families. The models considered in this paper are special cases of those considered by Sundberg, except for difficulties involving the possibility that some $p_i(\boldsymbol{\pi})$ or π_h may be 0. Equation (1.10) and the asymptotic normality results in Section 3 are consistent with those of Sundberg whenever the $p_i(\boldsymbol{\pi})$ and π_h are all positive; however, Sundberg has no results comparable to the convergence results of Sections 4 and 5.

2. Maximum likelihood equations. The maximum likelihood estimate $\hat{\boldsymbol{\pi}}$ is obtained by maximization of the log-likelihood kernel

$$l(\mathbf{n}^*, \boldsymbol{\pi}) = \sum_{j \in J} n_j{}^* \log p_j{}^*(\boldsymbol{\pi}) \tag{2.1}$$

for $\boldsymbol{\pi} \in \Theta_+$, where the convention $0 \log 0 = 0$ is used. Since $l(\mathbf{n}^*, \cdot)$ is continuous and bounded above by 0 and Θ_+ is compact, at least one maximum $\hat{\boldsymbol{\pi}}$ of $l(\mathbf{n}^*, \cdot)$ must exist. This maximum satisfies the conditions expressed in Theorem 1.

THEOREM 1. *Let $\hat{\boldsymbol{\pi}}$ be a maximum likelihood estimate of $\boldsymbol{\pi}$. Then* (1.10) *and* (1.11) *are satisfied by some $\mathbf{b} \in \Omega^\perp$, the orthogonal complement of $\Omega = \{\mathbf{z} - \mathbf{w} : \mathbf{z}, \mathbf{w} \in \Theta\}$.*

PROOF. The result is obtained by use of the Kuhn–Tucker lemma (see Zangwill (1969, pages 40–41)). To apply this lemma, the differential $dl_{\hat{\pi}}(n^*, \cdot)$ of $l(n^*, \cdot)$ at $\hat{\boldsymbol{\pi}}$ is obtained.

To find $dl_{\hat{\pi}}(\mathbf{n}^*, \cdot)$, note that if $\hat{\pi}_h > 0$, then (1.1), (1.8), (1.9) and (2.1) imply that

$$\frac{\partial}{\partial \pi_h} l(\mathbf{n}^*, \hat{\boldsymbol{\pi}}) = m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*)/\hat{\pi}_h , \tag{2.2}$$

while if $\hat{\pi}_h = 0$, then

$$\frac{\partial}{\partial \pi_h} l(\mathbf{n}^*, \hat{\boldsymbol{\pi}}) = \frac{\partial}{\partial \pi_h} m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) . \tag{2.3}$$

Thus if $H_1 = \{h \in H : \hat{\pi}_h > 0\}$ and $H_2 = H - H_1$, then

$$dl_{\hat{\pi}}(\mathbf{n}^*, \delta) = \sum_{h \in H_1} \delta_h m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*)/\hat{\pi}_h + \sum_{h \in H_2} \delta_h \frac{\partial}{\partial \pi_h} m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) . \tag{2.4}$$

Since $l(\mathbf{n}^*, \boldsymbol{\pi})$ is maximized by $\hat{\boldsymbol{\pi}}$ for $\boldsymbol{\pi} \in \Theta$ such that $\pi_h \geqq 0$, $h \in H$, the Kuhn–Tucker lemma implies that for some $\lambda_h \geqq 0$,

$$dl_{\hat{\pi}}(\mathbf{n}^*, \boldsymbol{\delta}) + (\boldsymbol{\lambda}, \boldsymbol{\delta}) = 0 , \qquad \boldsymbol{\delta} \in \Omega , \tag{2.5}$$

and

$$\lambda_h \hat{\pi}_h = 0 , \qquad h \in H . \tag{2.6}$$

Let

$$\begin{aligned} b_h &= m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*)/\hat{\pi}_h , \quad h \in H_1 , \\ &= \lambda_h + \frac{\partial}{\partial \pi_h} m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) , \quad h \in H_2 . \end{aligned} \tag{2.7}$$

By (2.4), (2.5) and (2.6), $\mathbf{b} \in \Omega^\perp$. Since $m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) = 0$ and $\lambda_h \geqq 0$ if $\hat{\pi}_h = 0$, (2.7) implies (1.10) and (1.11). □

Formulas (1.10) and (1.11) simplify in the direct observation case in which $I = J$ and $J_i = \{i\}$ for $i \in I$, for $\mathbf{n}^* = \mathbf{n}$ and

$$m_h(\boldsymbol{\pi} \mid \mathbf{n}^*) = \sum_{i \in I} n_i c(h, i) = y_h . \tag{2.8}$$

Thus (1.10) reduces to

$$y_h = b_h \hat{\pi}_h \tag{2.9}$$

and (1.11) reduces to the condition $b_h \geqq 0$ if $\hat{\pi}_h = 0$.

In the examples considered in this paper, it is helpful to employ the following corollaries to Theorem 1.

COROLLARY 1. *Assume that H_g, $g \in G$, are disjoint nonempty sets with union H, and assume that*

$$\Theta = \{\boldsymbol{\pi} \in R^H : \textstyle\sum_{h \in H_g} \pi_h = 1, g \in G\} .$$

Then any maximum likelihood estimate $\hat{\boldsymbol{\pi}}$ *of* $\boldsymbol{\pi}$ *satisfies the equations*

$$m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) = \hat{\pi}_h \textstyle\sum_{h' \in H_g} m_{h'}(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) , \qquad h \in H_g,\ g \in G , \tag{2.10}$$

$$\frac{\partial}{\partial \pi_h} m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) \leqq \textstyle\sum_{h' \in H_g} m_{h'}(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) , \qquad h \in H_g,\ g \in G,\ \hat{\pi}_h = 0 . \tag{2.11}$$

PROOF. Observe that $\Omega = \{\mathbf{x} \in R^H : \sum_{h \in H_g} x_h = 0,\ g \in G\}$ and $\Omega^\perp$ consists of $\mathbf{x} \in R^H$ such that for some $\mathbf{e} \in R^G$, $x_h = e_g$, $h \in H_g$, $g \in G$. By Theorem 1, for some $\mathbf{e} \in R^G$

$$m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) = e_g \hat{\pi}_h , \qquad h \in H_g,\ g \in G , \tag{2.12}$$

and

$$\frac{\partial}{\partial \pi_h} m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) \leqq e_g , \qquad \hat{\pi}_h = 0,\ h \in H_g,\ g \in G . \tag{2.13}$$

Addition of both sides of (2.13) over $h \in H_g$ shows that

$$e_g = \textstyle\sum_{h \in H_g} m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) . \tag{2.14}$$

Given (2.12), (2.13) and (2.14), (2.10) and (2.11) follow. □

COROLLARY 2. *Assume that* H_g, $g \in G$, *are nonempty disjoint sets with union* H *and for some* $N \subset H$, *the sets* B_c, $c \in C$, *are nonempty disjoint sets with union* $H - N$. *For each* $c \in C$, *assume that there exists a* $G_c \subset G$ *and a* $f_c > 0$ *such that* $H_g \cap B_c$ *has* f_c *elements if* $g \in G_c$ *and* $H_g \cap B_c$ *is empty if* $g \in G - G_c$. *Let* Θ *consist of* $\boldsymbol{\pi} \in R^H$ *such that*

$$\begin{aligned} \textstyle\sum_{h \in H_g} \pi_h &= 1 , && g \in G , \\ \pi_h &= \pi_{h'} , && h, h' \in B_c,\ c \in C , \\ \pi_h &= 0 , && h \in N . \end{aligned}$$

Let $\hat{\boldsymbol{\pi}}$ *be a maximum likelihood estimate of* $\boldsymbol{\pi}$. *Then* $\hat{\pi}_h = 0$ *for* $h \in N$, *and for* $h \in B_c$, $c \in C$,

$$\textstyle\sum_{h' \in B_c} m_{h'}(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) = f_c \hat{\pi}_h \sum_{g \in G_c} \sum_{h' \in H_g} m_{h'}(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) . \tag{2.15}$$

If $h \in B_c$, $c \in C$, *and* $\hat{\pi}_h = 0$, *then*

$$\sum_{h' \in B_c} \frac{\partial}{\partial \pi_{h'}} m_{h'}(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) \leqq f_c \sum_{g \in G_c} \sum_{h' \in H_g} m_{h'}(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) . \tag{2.16}$$

PROOF. Note that Ω consists $\mathbf{x} \in R^H$ such that

$$\begin{aligned} \textstyle\sum_{h \in H_g} x_h &= 0 , && g \in G , \\ x_h &= x_{h'} , && h, h' \in B_c,\ c \in C , \\ x_h &= 0 , && h \in N , \end{aligned}$$

and $\Omega^\perp$ consists of $\mathbf{x} \in R^H$ such that for some $\boldsymbol{\alpha} \in R^G$, $\boldsymbol{\beta} \in R^H$, and $\boldsymbol{\gamma} \in R^H$, $x_h = \alpha_g + \beta_h + \gamma_h$, $h \in H_g$, $g \in G$,

$$\beta_h = 0 , \qquad h \notin N , \tag{2.17}$$

$$\textstyle\sum_{h \in B_c} \gamma_h = 0 , \qquad c \in C . \tag{2.18}$$

By Theorem 1, there exist $\boldsymbol{\alpha} \in R^G$, $\boldsymbol{\beta} \in R^H$, and $\boldsymbol{\gamma} \in R^H$ such that (2.17) and (2.18) hold and

$$m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) = (\alpha_g + \beta_h + \gamma_h)\hat{\pi}_h , \qquad h \in H_g,\ g \in G , \tag{2.19}$$

$$\frac{\partial}{\partial \pi_h} m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) \leqq \alpha_g + \beta_h + \gamma_h , \qquad \hat{\pi}_h = 0,\ h \in H_g,\ g \in G . \tag{2.20}$$

To verify (2.15) and (2.16), note that for $c \in C$,

$$\begin{aligned} \textstyle\sum_{g \in G_c} \sum_{h \in H_g} m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) &= \textstyle\sum_{g \in G_c} \alpha_g + \sum_{g \in G_c} \sum_{h \in H_g} \gamma_h \hat{\pi}_h \\ &= \textstyle\sum_{g \in G_c} \alpha_g + \sum_{c \in C} \sum_{h \in B_c} \gamma_h \hat{\pi}_h \\ &= \textstyle\sum_{g \in G_c} \alpha_g , \end{aligned} \tag{2.21}$$

$$\textstyle\sum_{h' \in B_c} m_{h'}(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) = \hat{\pi}_h f_c \sum_{g \in G_c} \alpha_g , \quad h \in B_c , \tag{2.22}$$

and

$$\textstyle\sum_{h \in B_c} m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) \leqq f_c \sum_{g \in G_c} \alpha_g , \qquad \pi_h = 0 \quad \text{for} \quad h \in B_c . \tag{2.23}$$ □

2.1. *Examples of maximum likelihood equations.* In this section, the general results of Theorem 1 and Corollary 1 are applied to the examples from genetics and latent-structure analysis.

EXAMPLE 1. *The general model for estimation of gene frequencies* (*continued*). In this example, application of Corollary 1 is rather straightforward. A modest simplification of (2.10) and (2.11) is available since for each locus g,

$$\textstyle\sum_{h \in H_g} m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) ,$$

the conditional expected number of alleles $h \in H_g$ in the sample, must always equal $2N$. This result holds since each subject has 2 alleles from H_g at each locus. Consequently, (2.10) reduces to (1.12), and one has the added requirement

$$\frac{\partial}{\partial \pi_h} m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) \leqq 2N \qquad \text{if} \quad \hat{\pi}_h = 0 .$$

The term gene counting equation is applied to (1.12) since the left-hand side is the estimated proportion of genes in the population at locus g which have allele h, while the right-hand side is the estimated proportion of genes in the sample at locus g which have allele h.

EXAMPLE 2. *The two-loci model* (*continued*). In this model, it suffices to consider equations for $\hat{\pi}_{T'}$ and $\hat{\pi}_T$. The remaining equations follow from the relationships

$$\hat{\pi}_{t'} = 1 - \hat{\pi}_{T'}$$

and

$$\hat{\pi}_t = 1 - \hat{\pi}_T .$$

From Table 1, it follows that a subject with phenotype 1 has two alleles T', a subject with phenotype 2 has one such allele, and a subject with phenotype 3 has one allele T' with probability $1 - \hat\pi_{t'}^2\hat\pi_T^2/p_3(\hat\pi)$. The allele T' is not observed in phenotypes 4 and 5. Consequently, the gene counting equation for T' is

$$(2.24)\qquad \hat\pi_{T'} = \frac{1}{2N}\{2n_1^* + n_2^* + n_3^*[1 - \hat\pi_{t'}^2\hat\pi_T^2/p_3(\hat{\boldsymbol\pi})]\}.$$

Similar arguments show that

$$(2.25)\qquad \hat\pi_T = \frac{1}{2N}\{2n_1^*\hat\pi_T + 2n_2^* + n_3^*[1 - 2\hat\pi_{T'}\hat\pi_{t'}\hat\pi_t/p_3(\hat{\boldsymbol\pi})] + n_4^*\}.$$

EXAMPLE 3. *Latent-structure analysis (continued).* The maximum likelihood exuations depend on the choice of Θ, but it is always the case that for some $\mathbf{b} \in \Omega^\perp$,

$$(2.26)\qquad \begin{aligned} m_{\mathbf{k}}(\hat{\boldsymbol\pi} \mid \mathbf{n}^*) &= \textstyle\sum_{\mathbf{j}\in J} [p_{\mathbf{kj}}(\hat{\boldsymbol\pi})/p_{\mathbf{j}}^*(\hat{\boldsymbol\pi})]n_{\mathbf{j}}^* \\ &= b_{\mathbf{k}}\hat\pi_{\mathbf{k}}, \quad \mathbf{k} \in K, \end{aligned}$$

and

$$(2.27)\qquad \begin{aligned} m_{\mathbf{k}ug}(\hat{\boldsymbol\pi} \mid \mathbf{n}^*) &= \textstyle\sum_{\mathbf{j}\in J; j(u)=g} [p_{\mathbf{kj}}(\hat{\boldsymbol\pi})/p_j^*(\hat{\boldsymbol\pi})]n_{\mathbf{j}}^* \\ &= b_{\mathbf{k}ug}\hat\pi_{\mathbf{k}ug}, \quad \mathbf{k} \in K, \quad g \in L_u, \quad u \in U. \end{aligned}$$

In the classical latent-structure model with

$$\Theta = \Theta' = \{\boldsymbol\pi \in R^H: \textstyle\sum_{\mathbf{k}\in K} \pi_{\mathbf{k}} = 1, \sum_{g\in L_u} \pi_{\mathbf{k}ug} = 1, \mathbf{k} \in K, u \in U\},$$

Corollary 1, (2.26) and (2.27) imply that

$$(2.28)\qquad \hat\pi_{\mathbf{k}} = N^{-1}m_{\mathbf{k}}(\hat{\boldsymbol\pi} \mid \mathbf{n}^*)$$

and

$$(2.29)\qquad m_{\mathbf{k}ug}(\hat{\boldsymbol\pi} \mid \mathbf{n}^*) = \hat\pi_{\mathbf{k}ug}m_{\mathbf{k}}(\hat{\boldsymbol\pi} \mid \mathbf{n}^*).$$

If $m_{\mathbf{k}}(\hat{\boldsymbol\pi} \mid \mathbf{n}^*) > 0$,

$$(2.30)\qquad \hat\pi_{\mathbf{k}ug} = m_{\mathbf{k}ug}(\hat{\boldsymbol\pi} \mid \mathbf{n}^*)/m_{\mathbf{k}}(\hat{\boldsymbol\pi} \mid \mathbf{n}^*).$$

Note that in (2.28), the left-hand side is the estimated proportion of population members with $\mathbf{X} = \mathbf{k}$, and the right-hand side is the estimated proportion of sample member with $\mathbf{X} = \mathbf{k}$. In (2.30), the left-hand side is the estimated proportion of population members with $\mathbf{X} = \mathbf{k}$ for whom $A(u) = g$, and the right-hand side is the estimated proportion of sample members with $\mathbf{X} = \mathbf{k}$ for whom $A(u) = g$.

A more complex case with simple maximum likelihood equations is available if for each $u \in U$, K has a partition into disjoint nonempty sets D_{uw}, $w \in W_u$. Then one may consider a model with Θ consisting of $\pi \in \Theta'$ such that

$$\pi_{\mathbf{k}ug} = \pi_{\mathbf{k}'ug}, \qquad \mathbf{k}, \mathbf{k}' \in D_{uw},\ w \in W_u,\ g \in L_u,\ u \in U.$$

Corollary 2 then implies that (2.28) holds and if $g \in L_u$, $\mathbf{k} \in D_{uw}$, $w \in W_u$, and $u \in U$, then

$$(2.31)\qquad \textstyle\sum_{\mathbf{k}'\in D_{uw}} m_{\mathbf{k}'ug}(\hat{\boldsymbol\pi} \mid \mathbf{n}^*) = \hat\pi_{\mathbf{k}ug} \sum_{\mathbf{k}'\in D_{uw}} m_{\mathbf{k}'}(\hat{\boldsymbol\pi} \mid \mathbf{n}^*).$$

If

$$\sum_{\mathbf{k}' \in D_{uw}} m_{\mathbf{k}'}(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) > 0 ,$$

then

$$\hat{\pi}_{\mathbf{k}ug} = \sum_{\mathbf{k}' \in D_{uw}} m_{\mathbf{k}'ug}(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) / \sum_{\mathbf{k}' \in D_{uw}} m_{\mathbf{k}'}(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) . \tag{2.32}$$

Since $\hat{\pi}_{\mathbf{k}ug} = \pi_{\mathbf{k}'ug}$ for $\mathbf{k}, \mathbf{k}' \in D_{uw}$, the left-hand side of (2.32) is the estimated proportion of population members with $\mathbf{X} \in D_{uw}$ for whom $A(u) = g$, while the right-hand side is the corresponding estimated proportion for the sample.

EXAMPLE 4. *The two-latent-variable model (continued).* In the model with two latent variables considered in Section 1, (2.28) holds and (2.31) applies with $W_u = \{1, 2\}$ for $1 \leqq u \leqq 4$, $D_{uw} = \{\langle w, 1\rangle, \langle w, 2\rangle\}$, $u = 1$ or $u = 3$, and $D_{uw} = \{\langle 1, w\rangle, \langle 2, w\rangle\}$, $u = 2$ or $u = 4$.

2.2. *Multiple solutions to the maximum likelihood equations.* Although any maximum likelihood estimate $\hat{\boldsymbol{\pi}}$ satisfies (1.10) and (1.11) for some $\mathbf{b} \in \Omega^{\perp}$, not all solutions $\hat{\boldsymbol{\pi}}$ of (1.10) and (1.11) need be maximum likelihood estimates. For example, in latent-structure analysis (Example 3), if $\Theta = \Theta'$, $\hat{\pi}_{\mathbf{k}} = x_{\mathbf{k}} > 0$ for $\mathbf{k} \in K$, and

$$\hat{\pi}_{\mathbf{k}ug} = N^{-1} \sum_{\mathbf{j} \in J; j(u) = g} n_{\mathbf{j}}^* > 0 , \qquad g \in L_u,\ u \in U,\ \mathbf{k} \in K ,$$

then (2.31) and (2.32) are satisfied. Condition (1.11) holds since $\hat{\pi}_h > 0$ for all $h \in H$. Nonetheless, $\hat{\boldsymbol{\pi}}$ generally is not a maximum likelihood estimate of $\boldsymbol{\pi}$. Instead, $\hat{\boldsymbol{\pi}}$ is a maximum likelihood estimate of $\boldsymbol{\pi}$ for a restricted model with $\Theta = \{\boldsymbol{\pi} \in R^H : \pi_{\mathbf{k}ug} = \pi_{\mathbf{k}'ug} : g \in L_u, u \in U, \mathbf{k}, \mathbf{k}' \in K\}$. Consequently, conditions are desired for verification that a $\hat{\boldsymbol{\pi}} \in \Theta_+$ which satisfies (1.10) and (1.11) for some $\mathbf{b} \in \Omega^{\perp}$ is a maximum likelihood estimate. The following theorem can be helpful in this regard.

THEOREM 2. *Suppose that for some $\hat{\boldsymbol{\pi}} \in \Theta_+$ and $\mathbf{b} \in \Omega^{\perp}$,* (1.10) *and* (1.11) *are satisfied. If $\hat{\boldsymbol{\pi}}$ is a maximum likelihood estimate of $\boldsymbol{\pi}$, then*

$$(\mathbf{z}, E(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*)\mathbf{z}) \geqq 0 , \qquad \mathbf{z} \in \Omega'(\hat{\boldsymbol{\pi}}) . \tag{2.33}$$

On the other hand, if strict inequality holds in (1.11) *whenever $\hat{\pi}_h = 0$, and strict inequality holds in* (2.33) *whenever $\mathbf{z} \neq \mathbf{0}$ and $\mathbf{z} \in \Omega'(\hat{\boldsymbol{\pi}})$, then $\hat{\boldsymbol{\pi}}$ is an isolated relative maximum of $l(\mathbf{n}^*, \cdot)$.*

If for all $\boldsymbol{\pi} \in \Theta_+$,

$$(\mathbf{z}, E(\boldsymbol{\pi} \mid \mathbf{n}^*)\mathbf{z}) \geqq 0 , \qquad \mathbf{z} \in \Omega'(\boldsymbol{\pi}) , \tag{2.34}$$

then $\hat{\boldsymbol{\pi}}$ is a maximum likelihood estimate. If $\hat{\boldsymbol{\pi}}$ is also any isolated relative maximum of $l(\mathbf{n}^*, \cdot)$, then $\hat{\boldsymbol{\pi}}$ is the unique maximum likelihood estimate of $\boldsymbol{\pi}$.

PROOF. These results are proven by evaluation of the second differential $d^2 l_{\boldsymbol{\pi}}(\mathbf{n}^*, \cdot, \cdot)$ of $l(\mathbf{n}^*, \cdot)$ at $\boldsymbol{\pi} \in \Theta_+$.

If $\pi_h > 0$ and $\pi_{h'} > 0$, then

$$\begin{aligned} \frac{\partial^2}{\partial \pi_h \, \partial \pi_{h'}} l(\mathbf{n}^*, \boldsymbol{\pi}) &= \pi_h^{-1} \pi_{h'}^{-1} [C_{hh'}(\boldsymbol{\pi} \mid \mathbf{n}^*) - B_{hh'}(\boldsymbol{\pi} \mid \mathbf{n}^*)] \\ &= -E_{hh'}(\boldsymbol{\pi} \mid \mathbf{n}^*) . \end{aligned} \tag{2.35}$$

Consequently, if $\mathbf{w}, \mathbf{z} \in \Omega'(\boldsymbol{\pi})$,

$$d^2 l_\pi(\mathbf{n}^*, \mathbf{w}, \mathbf{z}) = -(\mathbf{w}, E(\boldsymbol{\pi} \mid \mathbf{n}^*)\mathbf{z}) . \tag{2.36}$$

By the second-order necessary condition of Fiacco and McCormick (1968, page 25), if $\hat{\boldsymbol{\pi}}$ is a maximum of $l(\mathbf{n}^*, \cdot)$, then the left-hand side of (2.36) must be nonpositive whenever $\mathbf{w} = \mathbf{z} \in \Omega'(\boldsymbol{\pi})$. Thus (2.33) holds. By Fiacco and McCormick's (1968, page 30) second-order sufficient condition, $\hat{\boldsymbol{\pi}}$ is an isolated maximum if the left-hand side of (2.36) is negative for all $\mathbf{w} = \mathbf{z} \in \Omega'(\hat{\boldsymbol{\pi}})$, $\mathbf{z} \neq \mathbf{0}$, and if strict inequality holds in (1.11). The former condition holds if strict inequality is present in (2.33) for all $\mathbf{z} \in \Omega'(\boldsymbol{\pi})$, $\mathbf{z} \neq \mathbf{0}$.

If (2.34) holds, then (2.36) implies that $l(\mathbf{n}^*, \cdot)$ is concave. Since Θ_+ is convex, $\hat{\boldsymbol{\pi}}$ is a maximum of $l(\mathbf{n}^*, \cdot)$ (see Zangwill (1969, page 43)). Furthermore, the set of all maxima must be convex. If $\hat{\boldsymbol{\pi}}$ is also an isolated relative maximum, then $\hat{\boldsymbol{\pi}}$ must be the unique maximum of $l(\mathbf{n}^*, \cdot)$. □

By far the simplest application of this theorem occurs when $\mathbf{n}$ is directly observed, so that $J = I$ and $J_i = \{i\}$ for $i \in I$. Then $C(\boldsymbol{\pi} \mid \mathbf{n}^*) = 0$ for all $\boldsymbol{\pi} \in \Theta_+$. Consequently, if $\hat{\boldsymbol{\pi}} \in \Theta_+$ satisfies (2.8) for a $\mathbf{b} \in \Omega^\perp$ such that $b_h \geqq 0$ if $\hat{\pi}_h \leqq 0$, then $\hat{\boldsymbol{\pi}}$ is a maxmum likelihood estimate. If $b_h > 0$ when $\hat{\pi}_h = 0$ and $m_h(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*) > 0$ whenever $\hat{\pi}_h > 0$, then $\hat{\boldsymbol{\pi}}$ is the unique maximum likelihood estimate. Thus estimation of $\hat{\boldsymbol{\pi}}$ is relatively uncomplicated in the direct observation case. This observation is of interest in its own right in estimation of transition probabilities in Markov chains (Anderson and Goodman (1957)) or in factorial contingency table models involving hypotheses of independence, conditional independence, and equiprobability (Goodman (1970)). The observation will also be helpful in exploration of properties of the functional iteration algorithm of Section 4.

In general, it is relatively difficult to verify that a solution of (1.10) and (1.11) is definitely a maximum likelihood estimate, even if

$$(\mathbf{z}, B(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*)\mathbf{z}) > (\mathbf{z}, C(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*)\mathbf{z}) , \qquad \mathbf{z} \in \Omega'(\hat{\boldsymbol{\pi}}),\ \mathbf{z} \neq \mathbf{0} , \tag{2.37}$$

and strict inequality holds in (1.11) if $\hat{\pi}_h = 0$. However, this problem appears to cause few practical problems, particularly if the numerical techniques of Sections 4 and 5 are employed.

3. Large-sample behavior of maximum likelihood estimates. Discussion of large-sample properties of maximum likelihood estimates is complicated by the possibility of multiple solutions of (1.10) and (1.11) and by the possibility that some π_h may be 0. Despite these difficulties, the following theorem is available.

THEOREM 3. *Let* $N_k/N \to \tau_k$ *as* $N \to \infty$. *Then* (1.18) *and* (1.19) *hold for some* $\boldsymbol{\beta} \in \Omega^\perp$. *Assume that strict inequality holds in* (1.19) *if* $\pi_h = 0$. *Assume* (1.20) *holds for* $\boldsymbol{\mu}^* = \{\tau_k p_j^*(\boldsymbol{\pi}) : j \in A_k,\ 1 \leqq k \leqq K\}$. *Then there exists an open neighborhood* $M \in Q = \{\mathbf{x}^* \in R^J : x_j^* = 0 \text{ if } p_j^*(\mathbf{x}) = 0,\ x_j^* > 0 \text{ if } p_j^*(\mathbf{x}) > 0\}$ *of* $\boldsymbol{\mu}^*$ *and an open neighborhood* $O \subset \Theta$ *of* $\boldsymbol{\pi}$ *such that if* $N^{-1}\mathbf{n}^* \in M$, *then there exists a unique* $\hat{\boldsymbol{\pi}}(\mathbf{n}^*) \in O \cap \Theta_+$ *such that for some* $b(\mathbf{n}^*) \in \Omega^\perp$, (1.10) *and* (1.11) *are satisfied* *by* $\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}}(\mathbf{n}^*)$ *and* $\mathbf{b} = \mathbf{b}(\mathbf{n}^*)$. *For* $N^{-1}\mathbf{n}^* \in M$, $\hat{\boldsymbol{\pi}}$ *is the unique maximum of* $l(\mathbf{n}^*, \boldsymbol{\pi})$ *for* $\boldsymbol{\pi} \in O \cap \Theta_+$. *As* $N \to \infty$, $N^{\frac{1}{2}}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi})$ *converges to* $N(0, \Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)$ *in distribution, where* $\Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)$ *is defined by* (1.21).

If $\boldsymbol{\pi}$ *is the only* $\mathbf{x} \in \Theta_+$ *such that* $p_j^*(\mathbf{x}) = p_j^*(\boldsymbol{\pi})$ *for all* $j \in A_k$ *such that* $\tau_k > 0$, *then the probability approaches* 1 *that* $\hat{\boldsymbol{\pi}}$ *is the unique maximum likelihood estimate of* $\boldsymbol{\pi}$.

REMARKS. The theorem's conditions for asymptotic normality of $\hat{\boldsymbol{\pi}}(\mathbf{n}^*)$ hold if $\pi_h > 0$ for all $h \in H$, $N_k/N \to \tau_k$ as $N \to \infty$, and $E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)$ is positive definite. Since $E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)$ is invertible, it follows from basic properties of projections that $\Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*) = P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)[E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^{-1}$, where $P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)$ is the projection on $\Omega = \Omega'(\boldsymbol{\pi})$ with respect to $E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)$. It can be shown that the conditions of the theorem cannot hold if $p_j^*(\boldsymbol{\pi}) > 0$ for all $j \in J$, if $\pi_h = 0$ for some $h \in H$, and if for some $\mathbf{x} \in \Theta_+$, $x_h > 0$ for all $h \in H$.

PROOF. To verify that (1.18) and (1.19) hold for some $\boldsymbol{\beta} \in \Omega^\perp$, note that $l(\boldsymbol{\mu}^*, \mathbf{x})$ is maximized for $\mathbf{x} \in \Theta_+$ if $\mathbf{x} = \boldsymbol{\pi}$, for the information inequality (Rao (1973, page 58)) implies that

$$\begin{aligned} l(\boldsymbol{\mu}^*, \boldsymbol{\pi}) &= \textstyle\sum_{k=1}^{r} \tau_k \sum_{j \in A_k} p_j^*(\boldsymbol{\pi}) \log p_j^*(\boldsymbol{\pi}) \\ &\geqq \textstyle\sum_{k=1}^{r} \tau_k \sum_{j \in A_k} p_j^*(\boldsymbol{\pi}) \log p_j^*(\mathbf{x}) \\ &= l(\boldsymbol{\mu}^*, \mathbf{x}) , \quad \mathbf{x} \in \Theta_+ . \end{aligned} \tag{3.1}$$

The implicit function theorem (Loomis and Sternberg (1968, page 231)) is used to show that suitable open sets M and O exist. To apply this theorem, consider the system of equations

$$dl_{\hat{\boldsymbol{\pi}}}\left(\frac{1}{N}\mathbf{n}^*, \boldsymbol{\delta}\right) + (\boldsymbol{\lambda}, \boldsymbol{\delta}) = 0 , \qquad \boldsymbol{\delta} \in \Omega , \tag{3.2}$$

$$\lambda_h \hat{\pi}_h = 0 , \qquad h \in H . \tag{3.3}$$

If $(1/N)\mathbf{n}^* = \boldsymbol{\mu}^*$, then (3.2) and (3.3) have a solution with $\hat{\boldsymbol{\pi}} \in \Theta_+$, for one may let $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}$, $\lambda_h = 0$ if $\pi_h > 0$, and $\lambda_h = \zeta_h = \beta_h - (\partial/\partial\pi_h)m_h(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*) > 0$ if $\pi_h = 0$.

For the implicit function theorem to apply, the equations

$$d^2 l_\pi(\boldsymbol{\mu}^*, \boldsymbol{\eta}, \boldsymbol{\delta}) + (\boldsymbol{\nu}, \boldsymbol{\delta}) = 0 , \qquad \boldsymbol{\delta} \in \Omega , \tag{3.4}$$

$$\zeta_h \eta_h + \pi_h \nu_h = 0 , \qquad h \in H , \tag{3.5}$$

must have the unique solution $\boldsymbol{\eta} = \boldsymbol{\nu} = \mathbf{0}$ for $\boldsymbol{\eta} \in \Omega$, $\boldsymbol{\nu} \in \Gamma^\perp$, where Γ consists of $\mathbf{z} \in \Omega^\perp$ such that $z_h = 0$ if $\pi_h > 0$. To solve (3.4) and (3.5) under the assumptions of the theorem, note that $\zeta_h > 0$ if $\pi_h = 0$ and $\zeta_h = 0$ if $\pi_h > 0$. If $\pi_h = 0$, then (3.5) implies that $\eta_h = 0$. If $\pi_h > 0$, then (3.5) implies that $\nu_h = 0$. By (2.36) and (3.4),

$$d^2 l_\pi(\boldsymbol{\mu}^*, \boldsymbol{\eta}, \boldsymbol{\eta}) + (\boldsymbol{\nu}, \boldsymbol{\eta}) = (\boldsymbol{\eta}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\boldsymbol{\eta}) = 0 . \tag{3.6}$$

Since $\boldsymbol{\eta} \in \Omega'(\boldsymbol{\pi})$, $\boldsymbol{\eta} = \mathbf{0}$. Thus (3.4) implies that

$$(3.7) \qquad (\boldsymbol{\nu}, \boldsymbol{\delta}) = 0 , \qquad \boldsymbol{\delta} \in \Omega .$$

Thus $\boldsymbol{\nu} \in \Gamma$. Since $\boldsymbol{\nu} \in \Gamma^{\perp}$, $\boldsymbol{\eta} = \boldsymbol{\nu} = \mathbf{0}$.

By the implicit function theorem, open neighborhoods M, O, and Λ exist such that $\boldsymbol{\mu}^* \in M \subset Q$, $\boldsymbol{\pi} \in O \subset \Theta$, and $\boldsymbol{\zeta} \in \Lambda \subset \boldsymbol{\zeta} + \Gamma^{\perp}$ and such that to each $(1/N)n^* \in M$ corresponds a unique $\hat{\boldsymbol{\lambda}}((1/N)\mathbf{n}^*) \in \Lambda$ and a unique $\hat{\boldsymbol{\pi}}((1/N)\mathbf{n}^*) \in 0$ such that (3.2) and (3.3) are satisfied by $\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}}((1/N)\mathbf{n}^*)$ and $\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\lambda}}((1/N)\mathbf{n}^*)$. The functions $\hat{\boldsymbol{\pi}}(\cdot)$ and $\hat{\boldsymbol{\lambda}}(\cdot)$ are continuously differentiable on M, and M can be chosen so that if $(1/N)\mathbf{n}^* \in M$, then

$$\hat{\pi}_h > 0 \quad \text{if} \quad \pi_h > 0 ,$$
$$\hat{\lambda}_h > 0 \quad \text{if} \quad \pi_h = 0 ,$$

and

$$(3.8) \qquad d^2 l_{\hat{\pi}}\left(\frac{1}{N}\mathbf{n}^*, \mathbf{v}, \mathbf{v}\right) < 0 \qquad \text{if} \quad \mathbf{v} \in \Omega'(\hat{\boldsymbol{\pi}}) , \quad \mathbf{v} \neq \mathbf{0} .$$

Given the choice of M, one may let $\Lambda = \boldsymbol{\zeta} + \Gamma^{\perp}$, for if $(1/N)\mathbf{n}^* \in M$, then (3.2) and (3.3) can hold for no more than one $\hat{\boldsymbol{\lambda}} \in \boldsymbol{\zeta} + \Gamma^{\perp}$. It should also be noted that if $(1/N)\mathbf{n}^* \in M$ and $\hat{\boldsymbol{\pi}} \in 0$, then (3.2) and (3.3) can only hold for a $\hat{\boldsymbol{\lambda}} \in R^H$ if they hold for some $\hat{\boldsymbol{\lambda}} \in \Lambda$. Thus if $(1/N)\mathbf{n}^* \in M$, then $\hat{\boldsymbol{\pi}}$ is the only relative maximum of $l((1/N)\mathbf{n}^*, \cdot)$ in 0.

For sufficiently small M and O, one has $l((1/N)\mathbf{n}^*, \hat{\boldsymbol{\pi}}) > l(\mathbf{w}^*, \mathbf{x})$ for $\mathbf{x} \in \partial O$, the boundary of O. Thus $\hat{\boldsymbol{\pi}}$ is the unique maximum of $l((1/N)\mathbf{n}^*, \cdot)$ in $O \cap \Theta_+$. Since $l(\mathbf{n}^*, \mathbf{x}) = Nl(N^{-1}\mathbf{n}^*, \mathbf{x})$ for $\mathbf{x} \in \Theta_+$, $\hat{\boldsymbol{\pi}}(\mathbf{n}^*) = \hat{\boldsymbol{\pi}}$ is the unique maximum of $l(\mathbf{n}^*, \cdot)$ in $O \cap \Theta_+$ and $\hat{\boldsymbol{\pi}}(\mathbf{n}^*)$ is the only element of $O \cap \Theta_+$ such that for some $\mathbf{b}(\mathbf{n}^*) = \mathbf{b} \in \Omega^{\perp}$, (1.10) and (1.11) hold.

To find the asymptotic distribution of $\hat{\boldsymbol{\pi}}$, $d\hat{\boldsymbol{\pi}}_{\mu^*}$ must be computed by use of (3.2), (3.3), and the implicit function theorem. Since $l(\cdot, \mathbf{x})$ is a linear function on R^J for each $\mathbf{x} \in \Theta_+$ and $\mathbf{w}^* \in R^J$, one has

$$(3.9) \qquad d^2 l_{\pi}(\boldsymbol{\mu}^*, d\hat{\pi}_{\mu^*}(\mathbf{w}^*), \boldsymbol{\delta}) + dl_{\pi}(\mathbf{w}^*, \boldsymbol{\delta}) + (d\hat{\boldsymbol{\lambda}}_{\mu^*}(\mathbf{w}^*), \boldsymbol{\delta}) = 0 , \qquad \boldsymbol{\delta} \in \Omega ,$$

$$(3.10) \qquad \zeta_h \, d\hat{\pi}_{h,\mu^*}(\mathbf{w}^*) + \pi_h \, d\hat{\lambda}_{h,\mu^*}(\mathbf{w}^*) = 0 , \qquad h \in H .$$

Given (2.12) and (2.36), it follows that if $\boldsymbol{\delta} \in \Omega'(\boldsymbol{\pi})$, then

$$(3.11) \qquad -(\boldsymbol{\delta}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*) \, d\hat{\boldsymbol{\pi}}_{\mu^*}(\mathbf{w}^*)) + (\boldsymbol{\delta}, \Pi^-(\boldsymbol{\pi})\mathbf{m}(\boldsymbol{\pi} \mid \mathbf{w}^*)) = 0 .$$

Equation (3.11) implies that

$$(3.12) \qquad d\hat{\boldsymbol{\pi}}_{\mu^*} = \Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\Pi^-(\boldsymbol{\pi})\mathbf{m}(\boldsymbol{\pi} \mid \cdot) .$$

To verify (3.12), note that by Halmos (1958, page 80), $[P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^A$ is a projection on R^H with range $\Delta^{\perp}$ and null space $[\Omega'(\boldsymbol{\pi})]^{\perp}$, where

$$\Delta = \{\mathbf{v} \in R^H : (\mathbf{v}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\mathbf{x}) = 0 \;\; \forall \, \mathbf{x} \in \Omega'(\boldsymbol{\pi})\} .$$

If $\mathbf{v} \in R^H$, then

$$(3.13)\quad \begin{aligned}&(\boldsymbol{\delta}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\mathbf{v})\\ &\quad = (\boldsymbol{\delta}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)[E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^-[P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^A\mathbf{v})\\ &\quad = (\{E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)[E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^-\}^A\boldsymbol{\delta}, [P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^A\mathbf{v})\\ &\quad = (\boldsymbol{\delta}, [P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^A\mathbf{v})\\ &\qquad - (\boldsymbol{\delta} - \{E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)[E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^-\}^A\boldsymbol{\delta}, [P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^A\mathbf{v})\,.\end{aligned}$$

If $\mathbf{x} \in \Omega'(\boldsymbol{\pi})$, then

$$(3.14)\quad \begin{aligned}&(\{E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)[E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^-\}^A\boldsymbol{\delta}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\mathbf{x})\\ &\quad = (\boldsymbol{\delta}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)[E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^-E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\mathbf{x})\\ &\quad = (\boldsymbol{\delta}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\mathbf{x})\,.\end{aligned}$$

Thus $\boldsymbol{\delta} - \{E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)[E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^-\}^A\boldsymbol{\delta} \in \Delta$. Since $[P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^A$ has range $\Delta^\perp$,

$$(3.15)\quad \begin{aligned}(\boldsymbol{\delta}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\mathbf{v}) &= (\boldsymbol{\delta}, [P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^A\mathbf{v})\\ &= (P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\boldsymbol{\delta}, \mathbf{v})\\ &= (\boldsymbol{\delta}, \mathbf{v})\,.\end{aligned}$$

Consequently, (3.11) reduces to

$$(3.16)\quad (\boldsymbol{\delta}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)[d\hat{\boldsymbol{\pi}}_{\boldsymbol{\mu}^*}(\mathbf{w}^*) - \Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\Pi^-(\boldsymbol{\pi})\mathbf{m}(\boldsymbol{\pi} \mid \mathbf{w}^*)]) = 0\,, \qquad \mathbf{w}^* \in R^J\,.$$

Since $d\hat{\boldsymbol{\pi}}_{\boldsymbol{\mu}^*}$ has range $\Omega'(\boldsymbol{\pi})$, (3.12) follows.

Given (3.12), standard arguments from large-sample theory (see Rao (1973, pages 385–389)) may be used to show that

$$(3.17)\quad N^{\frac{1}{2}}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) = \Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\Pi^-(\boldsymbol{\pi})\mathbf{m}(\boldsymbol{\pi} \mid N^{-\frac{1}{2}}(\mathbf{n}^* - \mathbf{e}^*)) + \mathbf{o}\,,$$

where $\mathbf{o}$ converges to $\mathbf{0}$ in probability and

$$(3.18)\quad e_j^* = N_k p_j^*(\boldsymbol{\pi})\,, \qquad j \in A_k,\ 1 \leqq k \leqq K\,.$$

Since $N^{-\frac{1}{2}}(\mathbf{n}^* - \mathbf{e}^*)$ converges to $N(\mathbf{0}, S)$ in distribution, where

$$(3.19)\quad \begin{aligned}(\mathbf{x}^*, S\mathbf{w}^*) = \textstyle\sum_{k=1}^K \tau_k\{&\textstyle\sum_{j \in A_k} x_j^* w_j^* p_j^*(\boldsymbol{\pi})\\ &- [\textstyle\sum_{j \in A_k} x_j^* p_j^*(\boldsymbol{\pi})][\textstyle\sum_{j \in A_k} w_j^* p_j^*(\boldsymbol{\pi})]\}\,,\end{aligned}$$

$\Pi^-(\boldsymbol{\pi})\mathbf{m}(\boldsymbol{\pi} \mid N^{-\frac{1}{2}}(\mathbf{n}^* - \mathbf{e}^*))$ converges to $N(\mathbf{0}, V(\boldsymbol{\pi} \mid \boldsymbol{\tau}))$ in distribution, and $N^{\frac{1}{2}}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi})$ and $\Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\Pi^-(\boldsymbol{\pi})\mathbf{m}(\boldsymbol{\pi} \mid N^{-\frac{1}{2}}(\mathbf{n}^* - \mathbf{e}^*))$ both converge in distribution to $N(\mathbf{0}, \Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)V(\boldsymbol{\pi} \mid \boldsymbol{\tau})\Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*))$. Here for $\boldsymbol{\eta}, \boldsymbol{\delta} \in R^H$,

$$(3.20)\quad \begin{aligned}&(\boldsymbol{\eta}, V(\boldsymbol{\pi} \mid \boldsymbol{\tau})\boldsymbol{\delta})\\ &\quad = \textstyle\sum_{k=1}^K \tau_k\{\textstyle\sum_{j \in A_k} (\mathbf{w}_j(\boldsymbol{\pi}), \boldsymbol{\eta})(\mathbf{w}_j(\boldsymbol{\pi}), \boldsymbol{\delta})p_j^*(\boldsymbol{\pi})\\ &\qquad - [\textstyle\sum_{j \in A_k} (\mathbf{w}_j(\boldsymbol{\pi}), \boldsymbol{\eta})p_j^*(\boldsymbol{\pi})][\textstyle\sum_{j \in A_k} (\mathbf{w}_j(\boldsymbol{\pi}), \boldsymbol{\delta})p_j^*(\boldsymbol{\pi})]\}\,,\end{aligned}$$

and

$$(3.21)\quad \begin{aligned}w_{hj}(\boldsymbol{\pi}) &= \pi_h^{-1} \textstyle\sum_{i \in J_j} c(h, i)p_i(\boldsymbol{\pi})/p_j^*(\boldsymbol{\pi})\,,\\ &\qquad \pi_h > 0 \quad\text{and}\quad p_j^*(\boldsymbol{\pi}) > 0\,,\\ &= 0\,, \qquad \pi_h = 0 \quad\text{or}\quad p_j^*(\boldsymbol{\pi}) = 0\,.\end{aligned}$$

Since

$$\sum_{j \in A_k} p_j^*(\mathbf{x}) = 1\,, \qquad \mathbf{x} \in \Theta_+,\ 1 \leqq k \leqq K\,, \tag{3.22}$$

differentiation of both sides of (3.22) shows that

$$\sum_{j \in A_k} (\mathbf{w}_j(\mathbf{x}), \boldsymbol{\delta}) p_j^*(\mathbf{x}) = 0\,, \qquad \boldsymbol{\delta} \in \Omega'(\mathbf{x}),\ \mathbf{x} \in \Theta_+\,. \tag{3.23}$$

Similarly, differentiation of both sides of (3.22) at $\mathbf{x} = \boldsymbol{\pi}$ shows that

$$\sum_{k=1}^{r} \tau_k \sum_{j \in A_k} (\mathbf{w}_j, \boldsymbol{\delta})(\mathbf{w}_j, \boldsymbol{\eta}) p_j^*(\boldsymbol{\pi}) = (\boldsymbol{\delta}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\boldsymbol{\eta})\,, \qquad \boldsymbol{\delta}, \boldsymbol{\eta} \in \Omega'(\boldsymbol{\pi})\,. \tag{3.24}$$

Thus

$$(\boldsymbol{\eta}, V(\boldsymbol{\pi} \mid \boldsymbol{\tau})\boldsymbol{\delta}) = (\boldsymbol{\eta}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\boldsymbol{\delta})\,, \qquad \boldsymbol{\delta}, \boldsymbol{\eta} \in \Omega'(\boldsymbol{\pi})\,. \tag{3.25}$$

Given (3.25), it is possible to show that $V(\boldsymbol{\pi} \mid \boldsymbol{\tau})$ is a generalized inverse of $\Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)$. To do so, note that for $\mathbf{x} \in \Omega'(\boldsymbol{\pi})$,

$$\begin{aligned} (\mathbf{z}, [P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^A E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\mathbf{x}) &= (P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\mathbf{z}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\mathbf{x}) \\ &= (P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\mathbf{z}, V(\boldsymbol{\pi} \mid \boldsymbol{\tau})\mathbf{x}) \\ &= (\mathbf{z}, [P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^A V(\boldsymbol{\pi} \mid \boldsymbol{\tau})\mathbf{x})\,, \quad \mathbf{z} \in R^H\,. \end{aligned} \tag{3.26}$$

Thus $[P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^A E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\mathbf{x} = [P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^A V(\boldsymbol{\pi} \mid \boldsymbol{\tau})\mathbf{x}$. Furthermore, by Rao and Mitra (1971, pages 3 and 20),

$$\begin{aligned} &(\mathbf{z}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)[E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^-[P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^A V(\boldsymbol{\pi} \mid \boldsymbol{\tau})\mathbf{x}) \\ &\quad = (\mathbf{z}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)[E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^-[P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^A E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\mathbf{x}) \\ &\quad = (\mathbf{z}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)[E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)]^- E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)P(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\mathbf{x}) \\ &\quad = (\mathbf{z}, E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)\mathbf{x})\,, \quad \mathbf{z} \in \Omega'(\boldsymbol{\pi})\,. \end{aligned} \tag{3.27}$$

Thus $\Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)V(\boldsymbol{\pi} \mid \boldsymbol{\tau})\mathbf{x} = \mathbf{x}$. Hence $V(\boldsymbol{\pi} \mid \boldsymbol{\tau})$ is a generalized inverse of $\Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)$.

If $\boldsymbol{\pi}$ is the only $\mathbf{x} \in \Theta_+$ such that $p_j^*(\mathbf{x}) = p_j^*(\boldsymbol{\pi})$ for $j \in A_k$ such that $\tau_k > 0$, then since $\Theta_+ - O \cap \Theta_+$ is compact, there exists some $\varepsilon > 0$ such that

$$\sum \tau_k |p_j^*(\mathbf{x}) - p_j^*(\boldsymbol{\pi})| \geqq \varepsilon$$

if $\mathbf{x} \in \Theta_+ - O \cap \Theta_+$. If $\bar{\boldsymbol{\pi}}(\mathbf{n}^*)$ is an arbitrary maximum likelihood estimate of $\boldsymbol{\pi}$, then by Rao (1973, page 356), $\tau_k p_j^*(\bar{\boldsymbol{\pi}}(\mathbf{n}^*))$ converges in probability to $\tau_k p_j^*(\boldsymbol{\pi})$ for all $j \in A_k$ and k such that $1 \leqq k \leqq K$. Therefore, the probability approaches 1 that $N^{-1}\mathbf{n}^* \in M$ and $\bar{\boldsymbol{\pi}}(\mathbf{n}^*) \in O \cap \Theta_+$. Since $\hat{\boldsymbol{\pi}}$ is the unique maximum of $l(\mathbf{n}^*, \mathbf{x})$ for $\mathbf{x} \in O \cap \Theta_+$ if $N^{-1}\mathbf{n}^* \in M$, the probability approaches 1 that $\bar{\boldsymbol{\pi}}(\mathbf{n}^*) = \hat{\boldsymbol{\pi}}$, the unique maximum likelihood estimate of $\boldsymbol{\pi}$. □

Given Theorem 3, $\hat{\boldsymbol{\pi}}$ may be said to be asymptotically unbiased with asymptotic variance $\Sigma(\boldsymbol{\pi} \mid \mathbf{e}^*)$, where $\mathbf{e}^*$ is defined as in (3.18). It is easily verified that $N\Sigma(\boldsymbol{\pi} \mid \mathbf{e}^*) \to \Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)$. Various estimates of $\Sigma(\boldsymbol{\pi} \mid \mathbf{e}^*)$ are possible. A common choice is $\hat{\Sigma}(\boldsymbol{\pi} \mid \mathbf{e}^*) = \Sigma(\hat{\boldsymbol{\pi}} \mid \mathbf{e}^*)$, where

$$\hat{e}_j^* = N_k p_j^*(\hat{\boldsymbol{\pi}})\,, \qquad j \in A_k,\ 1 \leqq k \leqq K\,. \tag{3.28}$$

An alternative is $\Sigma(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*)$.

It may readily be shown that both $N\hat{\Sigma}(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*)$ and $N\hat{\Sigma}(\hat{\boldsymbol{\pi}} \mid \mathbf{e}^*)$ converge to $\Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)$ in probability. Thus if $\boldsymbol{\gamma} \in R^H$ and $0 < \alpha < 1$, an approximate level-$(1 - \alpha)$ confidence interval for $(\boldsymbol{\gamma}, \boldsymbol{\pi})$ is the interval

$$(\boldsymbol{\gamma}, \hat{\boldsymbol{\pi}}) \pm Z_{\alpha/2}(\boldsymbol{\gamma}, \hat{\Sigma}(\boldsymbol{\mu} \mid \mathbf{e}^*)\boldsymbol{\gamma})^{\frac{1}{2}},$$

where $Z_{\alpha/2}$ is the upper $\alpha/2$-point of the standard normal distribution. An alternative interval is

$$(\boldsymbol{\gamma}, \hat{\boldsymbol{\pi}}) \pm Z_{\alpha/2}(\boldsymbol{\gamma}, \hat{\Sigma}(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*)\boldsymbol{\gamma})^{\frac{1}{2}}.$$

3.1. *Alternative formulas and conditions.* If $\Omega'(\boldsymbol{\pi})$ has dimension q, W is a linear transformation from R^q onto $\Omega'(\boldsymbol{\pi})$, and $\boldsymbol{\zeta} \in \Omega'(\boldsymbol{\pi})$, then for some unique $\boldsymbol{\gamma}$, $\boldsymbol{\pi} = W\boldsymbol{\gamma} + \boldsymbol{\zeta}$. Under the conditions of Theorem 3, the probability approaches 1 that $\hat{\boldsymbol{\pi}} - \boldsymbol{\zeta} \in \Omega'(\boldsymbol{\pi})$ and $\hat{\boldsymbol{\pi}} = W\hat{\boldsymbol{\gamma}} + \boldsymbol{\zeta}$ for a unique $\hat{\boldsymbol{\gamma}} \in R^q$. Since

$$\begin{aligned}(3.29)\qquad \hat{\boldsymbol{\gamma}} &= [W^A E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)W]^{-1}W^A E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)(\hat{\boldsymbol{\pi}} - \boldsymbol{\zeta}), \\ &= \boldsymbol{\gamma} + [W^A E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)W]^{-1}W^A E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}),\end{aligned}$$

an elementary calculation shows that $N^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ converges in distribution to $N(\mathbf{0}, [W^A E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)W]^{-1})$. It also follows that $N^{\frac{1}{2}}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi})$ converges to $N(\mathbf{0}, W[W^A E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)W]^{-1}W^A)$ in distribution; that is,

$$(3.30)\qquad \Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*) = W[W^A E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)W]^{-1}W^A.$$

Alternative expressions can be obtained by noting that

$$W^A E(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)W = W^A V(\boldsymbol{\pi} \mid \boldsymbol{\tau})W.$$

Also note that $W^A V(\boldsymbol{\pi} \mid N)W$ is the covariance operator of $W^A \Pi^-(\boldsymbol{\pi})\underset{\cdot}{\mathbf{m}}(\boldsymbol{\pi} \mid \mathbf{n}^*)$.

Alternatives to (1.20) exist in Theorem 3. The following conditions are equivalent to (1.20):

(A) $(\mathbf{z}, V(\boldsymbol{\pi} \mid \boldsymbol{\tau})\mathbf{z}) > 0$ if $\mathbf{z} \in \Omega'(\boldsymbol{\pi})$, $\mathbf{z} \neq \mathbf{0}$;

(B) for some open sets N' and O' such that $\boldsymbol{\mu}^* \in N' \subset [0, \infty)^J$, $\boldsymbol{\pi} \in O' \subset \boldsymbol{\pi} + \Omega'(\boldsymbol{\pi})$, there exists a differentiable function Z from N' to O' such that if $\mathbf{x} \in O' \cap \Theta_+$ and

$$w_j^* = \tau_k p_j^*(\mathbf{x}), \qquad j \in A_k,\ 1 \leqq k \leqq K,$$

then

$$(3.31)\qquad Z(\mathbf{w}^*) = \mathbf{x}.$$

(C) If $\boldsymbol{\delta} \in \Omega'(\boldsymbol{\pi})$ and

$$(3.32)\qquad \tau_k p_j^*(\boldsymbol{\pi})(\mathbf{w}_j(\boldsymbol{\pi}), \boldsymbol{\delta}) = 0, \qquad j \in A_k,\ 1 \leqq k \leqq K,$$

then $\boldsymbol{\delta} = \mathbf{0}$.

Verification of the equivalence of (1.20) to (A) and (C) is easily accomplished by use of (3.21), (3.23), and (3.25). That (1.20) implies (B) follows from Theorem 3 by replacing Ω by $\Omega'(\boldsymbol{\pi})$. The resulting function $\hat{\boldsymbol{\pi}}$ is a possible Z in (B).

That (B) implies (C) follows since differentiation of (3.31) leads to the equation

$$dZ_{\mu^*}\{\tau_k(\mathbf{w}_j(\boldsymbol{\pi}), \mathbf{z})p_j^*(\mathbf{z}) : j \in A_k, 1 \leqq k \leqq K\} = \mathbf{z}\,, \qquad \mathbf{z} \in \Omega'(\boldsymbol{\pi})\,. \tag{3.33}$$

If $\mathbf{z} \neq \mathbf{0}$ and $\mathbf{z} \in \Omega'(\boldsymbol{\pi})$, then $\tau_k > 0$, $(\mathbf{w}_j(\boldsymbol{\pi}), \mathbf{z}) \neq 0$, and $p_j^*(\boldsymbol{\pi}) > 0$, for some k and some $j \in A_k$. Thus (C) holds.

3.2. *Applicability of Theorem* 3. The results of Theorem 3 need not apply to all genetics models of the type considered in Example 1 or to all latent structure models of the type considered in Example 3. Nonetheless, Theorem 3 can be used with the two-loci model of Example 2 or the two-latent-variable model of Example 4.

The two-loci model. In the two-loci model, it is known from the data that $\pi_{T'}$, $\pi_{t'}$, π_T, and π_t are all positive, for all phenotypes occur with positive frequency. Thus the conditions of Theorem 3 will be satisfied if $N \to \infty$ and a function Z and sets N' and O' are available which satisfy condition (B). Such a function Z can be constructed with $N' = (0, 1)^5$ and $O' = \{\boldsymbol{\pi} \in \Theta : \pi_h > 0\}$ by use of the formulas

$$\begin{aligned} Z_T'(\mathbf{w}^*) &= (w_1^*)^{\frac{1}{2}}\,, \\ Z_t'(\mathbf{w}^*) &= 1 - Z_{t'}(\mathbf{w}^*)\,, \\ Z_t(\mathbf{w}^*) &= (w_5^*)^{\frac{1}{2}}/Z_{t'}(\mathbf{w}^*)\,, \\ Z_T(\mathbf{w}^*) &= 1 - Z_t(\mathbf{w}^*)\,. \end{aligned}$$

Given this Z, it also follows that if $p_j^*(\mathbf{x}) = p_j^*(\boldsymbol{\pi})$ and $\mathbf{x} \in \theta_+$, then $\mathbf{x} = \boldsymbol{\pi}$. Thus the $\hat{\boldsymbol{\pi}}$ in Theorem 3 can be chosen as any maximum likelihood estimate of $\boldsymbol{\pi}$. The probability approaches 1 that the choice is uniquely defined.

Observe that

$$\begin{bmatrix} \pi_{T'} \\ \pi_{t'} \\ \pi_T \\ \pi_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \pi_{T'} \\ \pi_T \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

Thus the asymptotic covariance operator of $\hat{\boldsymbol{\pi}}$ is determined by the asymptotic covariance matrix of $\langle \hat{\pi}_{T'}, \hat{\pi}_T \rangle$, which is the inverse of the covariance matrix of

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} m_{T'}(\boldsymbol{\pi} \mid \mathbf{n}^*) \\ m_{t'}(\boldsymbol{\pi} \mid \mathbf{n}^*) \\ m_T(\boldsymbol{\pi} \mid \mathbf{n}^*) \\ m_{t'}(\boldsymbol{\pi} \mid \mathbf{n}^*) \end{bmatrix} = \begin{bmatrix} [m_{T'}(\boldsymbol{\pi} \mid \mathbf{n}^*) - 2N\pi_{T'}]/(\pi_{T'}\pi_{t'}) \\ [m_T(\boldsymbol{\pi} \mid \mathbf{n}^*) - 2N\pi_T]/(\pi_T\pi_t) \end{bmatrix}.$$

Here the relationships

$$m_{t'}(\boldsymbol{\pi} \mid \mathbf{n}^*) = 2N - m_{T'}(\boldsymbol{\pi} \mid \mathbf{n}^*)$$

and

$$m_t(\boldsymbol{\pi} \mid \mathbf{n}^*) = 2N - m_T(\boldsymbol{\pi} \mid \mathbf{n}^*)\,,$$

have been used. Given the equations

$$m_{T'}(\boldsymbol{\pi} \mid \mathbf{n}^*) = 2n_1^* + n_2^* + n_3^*[1 - \pi_{t'}^2\pi_T{}^2/p_3(\boldsymbol{\pi})]$$

and

$$m_T(\boldsymbol{\pi} \mid \mathbf{n}^*) = 2n_1^*\pi_T + 2n_2^* + n_3^*[1 - 2\pi_{T'}\pi_{t'}\pi_t/p_3(\boldsymbol{\pi})] + n_4^* ,$$

computation of the asymptotic covariance matrix of $\langle \hat{\pi}_{T'}, \hat{\pi}_T \rangle$ is straightforward. Details are omitted.

The two-latent variable example. In this example, conditions of Theorem 3 hold if $N \to \infty$ and if $\boldsymbol{\pi}$ satisfies the following conditions:

$$\pi_h > 0 \qquad \text{for all} \quad h \in H ,$$

$$\pi_{\langle 1,1\rangle}\pi_{\langle 2,2\rangle}/(\pi_{\langle 1,2\rangle}\pi_{\langle 2,1\rangle}) \neq 1 ,$$

$$\pi_{\mathbf{k}ug} \neq \pi_{\mathbf{k}'ug} \qquad \text{if} \quad 1 \leqq g \leqq 2 , \quad \mathbf{k}, \mathbf{k}' \in D_{uw} , \quad 1 \leqq w \leqq 2 , \quad 1 \leqq u \leqq 4$$

where $D_{uw} = \{\langle w, 1\rangle, \langle w, 2\rangle\}$, $u = 1$ or 3, and $D_{uw} = \{\langle 1, w\rangle, \langle 2, w\rangle\}$, $u = 2$ or 4. This claim may be verified by use of a determinantal estimate of $\boldsymbol{\pi}$ as the function Z in condition (B). This estimate is constructed in Goodman (1974).

The estimate $\hat{\boldsymbol{\pi}}$ cannot be a unique maximum likelihood estimate since $l(\mathbf{n}^*, \hat{\boldsymbol{\pi}}) = l(\mathbf{n}^*, \bar{\boldsymbol{\pi}})$ if

$$\bar{\boldsymbol{\pi}}_{\mathbf{k}} = \hat{\pi}_{\sigma(\mathbf{k})} ,$$

$$\bar{\boldsymbol{\pi}}_{\mathbf{k}ug} = \hat{\pi}_{\sigma(\mathbf{k})ug} ,$$

and

$$\sigma(\langle k(1), k(2)\rangle) = \langle \sigma_1(k(1)), \sigma_2(k(2))\rangle$$

for some permutations σ_1 and σ_2 of $\{1, 2\}$. The estimate $\hat{\boldsymbol{\pi}}$ can be defined as the maximum of $l(\mathbf{n}^*, \mathbf{x})$ for $\mathbf{x} \in \Theta_+$ such that $x_{\langle 1,1\rangle 11} > x_{\langle 2,1\rangle 11}$ and $x_{\langle 1,2\rangle 21} > x_{\langle 1,2\rangle 21}$. This restriction causes no difficulty since the labelling of the latent variable classes is completely arbitrary. Computation of $\Sigma(\boldsymbol{\pi} \mid \boldsymbol{\mu}^*)$ in this example is somewhat tedious, but involves no special difficulties. Consequently, details will be omitted.

3.3. *Testing goodness of fit.* The customary statistics for general tests of goodness of fit are the log likelihood ratio chi-square

$$L^2 = 2 \textstyle\sum_{k=1}^{r} \sum_{j \in I} n_j^* \log \{n_j^*/[N_k p_j^*(\hat{\boldsymbol{\pi}})]\} \tag{3.34}$$

and the Pearson chi-square

$$X^2 = \textstyle\sum_{k=1}^{r} \sum_{j \in J} \{n_j^* - N_k p_j^*(\hat{\boldsymbol{\pi}})\}^2/[N_k p_j^*(\hat{\boldsymbol{\pi}})] , \tag{3.35}$$

where it is assumed that $0 \log 0 = \frac{0}{0} = 0$. Under the conditions of Theorem 3, together with the added condition that $\tau_k > 0$ for $1 \leqq k \leqq K$, L^2 and X^2 are asymptotically equivalent, and both statistics converge in distribution to the X_s^2 distribution, where

$$s = c(J^*) - r - \dim \Omega'(\boldsymbol{\pi}) \tag{3.36}$$

and $c(J^*)$ is the number of elements in $J^* = \{j \in J\colon p_j^*(\boldsymbol{\pi}) > 0\}$. Statistical tests may in practice be made with an estimate $\hat{s}$ for s, where

$$\hat{s} = c(\{j \in J\colon p_j^*(\hat{\boldsymbol{\pi}}) > 0\}) - r - \dim \Omega'(\hat{\boldsymbol{\pi}}) . \tag{3.37}$$

As $N \to \infty$, $P\{\hat{s} = s\} \to 1$. Verification of results in the section involves nn special difficulties, so that proofs are omitted.

4. Solution of the likelihood equations by functional iteration. The simplest approach to the determination of a solution $\hat{\pi}$ which satisfies (1.10) for some $\mathbf{b} \in \Omega^{\perp}$ is to use an initial approximation $\boldsymbol{\pi}^{(0)} \in \Theta_+$ for $\hat{\boldsymbol{\pi}}$ to generate a sequence $\{\boldsymbol{\pi}^{(v)}\} \subset \Theta_+$ of approximations to $\hat{\boldsymbol{\pi}}$ by the equation

$$\boldsymbol{\pi}^{(v+1)} = \bar{\boldsymbol{\pi}}(f(\boldsymbol{\pi}^{(v)})) , \tag{4.1}$$

where

$$f_i(\mathbf{x}) = n_j^* p_i(\mathbf{x})/p_j^*(\mathbf{x}) , \qquad i \in J_j,\ j \in J , \tag{4.2}$$

and for $\mathbf{z} \in [0, \infty)^I$, $\bar{\boldsymbol{\pi}}(\mathbf{z})$ is any maximum of

$$\bar{l}(\mathbf{z}, \mathbf{x}) = \sum_{i \in I} z_i \log p_i(\mathbf{x}) \tag{4.3}$$

for $\mathbf{x} \in \Theta_+$. Versions of this algorithm have been developed by Ceppellini, Siniscalco and Smith (1955), Chen (1972), and Goodman (1974); however, a systematic investigation of convergence properties has been lacking.

In this section, it is shown that the algorithm, although often slow to converge, is very stable numerically and often very easy to apply. Results are illustrated through use of the examples from genetics and latent-structure analysis.

In the direct observation case with $I = J$ and $J_i = \{i\}$, $f(\mathbf{x}) = \mathbf{n}$ and $\boldsymbol{\pi}^{(1)}$ is a maximum likelihood estimate of $\boldsymbol{\pi}$. More generally, $\{\boldsymbol{\pi}^{(v)}\}$ has the convergence properties discussed in the following theorem.

THEOREM 4. *Suppose* $\boldsymbol{\pi}^{(0)} \in \Theta_+$ *and* $l(\mathbf{n}^*, \boldsymbol{\pi}^{(0)}) > -\infty$. *Then the sequence* $\{\boldsymbol{\pi}^{(v)}\} \subset \Theta_+$ *defined by* (4.1) *is such that for* $v \geqq 0$, *either*

$$l(\mathbf{n}^*, \alpha\boldsymbol{\pi}^{(v)} + (1 - \alpha)\boldsymbol{\pi}^{(v+1)}) > l(\mathbf{n}^*, \boldsymbol{\pi}^{(v)}) , \qquad 0 \leqq \alpha < 1 , \tag{4.4}$$

or for some $\mathbf{b} \in \Omega^{\perp}$,

$$m_h(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*) = b_h \pi_h^{(v)} , \qquad h \in H . \tag{4.5}$$

Every subsequence of $\{\boldsymbol{\pi}^{(v)}\}$ *contains a convergent subsequence, and if* $\hat{\boldsymbol{\pi}}$ *is a limit point of* $\{\boldsymbol{\pi}^{(v)}\}$, *then* $\hat{\boldsymbol{\pi}}$ *satisfies* (1.10) *for some* $\mathbf{b} \in \Omega^{\perp}$.

PROOF. To verify (4.4), note that for any $\mathbf{x}, \mathbf{z} \in \Theta_+$,

$$l(\mathbf{n}^*, \mathbf{x}) = \bar{l}(f(\mathbf{z}), \mathbf{x}) + q(f(\mathbf{z}), \mathbf{x}) , \tag{4.6}$$

where for $\mathbf{u} \in [0, \infty)^I$,

$$q(\mathbf{u}, \mathbf{x}) = -\sum_{j \in J} \sum_{i \in J_j} u_i \log [p_i(\mathbf{x})/p_j^*(\mathbf{x})] . \tag{4.7}$$

Note that $\bar{l}(\mathbf{n}, \cdot)$ is the log likelihood kernel if $\mathbf{n}$ is directly observed. As shown in Section 2, $\bar{l}(f(\mathbf{z}), \cdot)$ is concave for any $\mathbf{z} \in \Theta_+$. Since $\bar{l}(f(\boldsymbol{\pi}^{(v)}, \boldsymbol{\pi}^{(v+1)}) \geqq \bar{l}(f(\boldsymbol{\pi}^{(v)}), \boldsymbol{\pi}^{(v)})$,

$$\bar{l}(f(\boldsymbol{\pi}^{(v)}), \alpha\boldsymbol{\pi}^{(v)} + (1 - \alpha)\boldsymbol{\pi}^{(v+1)}) \geqq \bar{l}(f(\boldsymbol{\pi}^{(v)}), \boldsymbol{\pi}^{(v)}) , \qquad 0 \leqq \alpha < 1 . \tag{4.8}$$

Furthermore, strict inequality holds in (4.8) unless $\boldsymbol{\pi}^{(v)}$ is a maximum of $\bar{l}(f(\boldsymbol{\pi}^{(v)}), \cdot)$. In this latter case, Theorem 1 implies that (4.5) holds for some $\mathbf{b} \in \Omega^{\perp}$. For any $\mathbf{x} \in \Theta_{+}$, the information inequality implies that

$$\begin{aligned} & q(f(\boldsymbol{\pi}^{(v)}), \mathbf{x}) \\ (4.9) \qquad & = -\sum_{j \in J} n_j^* \sum_{i \in J_j} [p_i(\boldsymbol{\pi}^{(v)})/p_j^*(\boldsymbol{\pi}^{(v)})] \log [p_i(\mathbf{x})/p_j^*(\mathbf{x})] \\ & \geqq -\sum_{j \in J} n_j^* \sum_{i \in J_j} [p_i(\boldsymbol{\pi}^{(v)})/p_j^*(\boldsymbol{\pi}^{(v)})] \log [p_i(\boldsymbol{\pi}^{(v)})/p_j^*(\boldsymbol{\pi}^{(v)})] \\ & = q(f(\boldsymbol{\pi}^{(v)}), \boldsymbol{\pi}^{(v)}) . \end{aligned}$$

Given (4.6), it follows that (4.4) holds unless (4.5) is satisfied for some $\mathbf{b} \in \Omega^{\perp}$.

Since Θ_{+} is compact, every subsequence of $\{\boldsymbol{\pi}^{(v)}\}$ contains a convergent subsequence. If $A(\mathbf{z})$ is the set of $\mathbf{x} \in \Theta_{+}$ which maximize $\bar{l}(\mathbf{z}, \cdot)$, then by Zangwill (1969, page 156), $A(\mathbf{z})$ is a closed point-to-set map from $[0, \infty)^I$ to Θ_{+}; that is, if $\mathbf{z}^{(t)} \to \mathbf{z}$, $\mathbf{x}^{(t)} \to \mathbf{x}$, and $\mathbf{x}^{(t)} \in A(\mathbf{z}^{(t)})$ for $t \geqq 0$, then $\mathbf{x} \in A(\mathbf{z})$. Provided that $l(\mathbf{n}^*, \mathbf{x}) > -\infty$, f is continuous at $\mathbf{x}$. By Zangwill (1969, page 96, Corollary 4.2.2 and page 91, Convergence Theorem A), every convergent subsequence of $\{\boldsymbol{\pi}^{(v)}\}$ has a limit $\hat{\boldsymbol{\pi}}$ such that (1.10) holds for some $\mathbf{b} \in \Omega^{\perp}$.

The functional iteration algorithm need not converge to a maximum likelihood estimate of $\boldsymbol{\pi}$; however, (4.4) implies that if $\boldsymbol{\pi}^{(0)}$ is sufficiently close to an isolated maximum $\hat{\boldsymbol{\pi}}$ of $l(\mathbf{n}^*, \cdot)$, then $\boldsymbol{\pi}^{(v)} \to \hat{\boldsymbol{\pi}}$. If the conditions of Theorem 3 are satisfied and $\boldsymbol{\pi}^{(0)} = T(\mathbf{n}^*)$, where $T(\mathbf{n}^*)$ converges to $\boldsymbol{\pi}$ in probability, then the probability approaches 1 that $\boldsymbol{\pi}^{(v)} \to \hat{\boldsymbol{\pi}}(\mathbf{n}^*)$. Given the results of Section 3.2, these remarks apply to both the two-loci example from genetics and the two-latent-variable example.

Implementation of the functional iteration algorithm is rather easy if the conditions of Corollaries 1 or 2 are satified. If the conditions of Corollary 1 are satisfied, then given $\boldsymbol{\pi}^{(0)}$, $\{\boldsymbol{\pi}^{(v)}\}$ is any sequence in Θ_{+} satisfying the equation

$$(4.10) \qquad m_h(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*) = \pi_h^{(v+1)} \sum_{h' \in H_g} m_{h'}(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*) , \qquad h \in H_g,\ g \in G,\ v \geqq 0 .$$

Similarly, if the conditions of Corollary 2 are satisfied, then given $\boldsymbol{\pi}^{(0)}$, $\{\boldsymbol{\pi}^{(v)}\}$ is any sequence in Θ_{+} satisfying the equations $\pi_h^{(v)} = 0$, $h \in N$, $v \geqq 0$ and

$$(4.11) \qquad \sum_{h' \in B_c} m_{h'}(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*) = f_c \pi_h^{(v+1)} \sum_{g \in G_c} \sum_{h' \in H_g} m_{h'}(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*) , \qquad h \in B_c,\ c \in C,\ v \geqq 0 .$$

In the general genetics model, the functional iteration algorithm reduces to the gene counting algorithm

$$(4.12) \qquad \pi_h^{(v+1)} = \frac{1}{2N} m_h(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*)$$

of Ceppellini, Siniscalco and Smith (1955).

In latent-structure analysis models of the type considered in Goodman (1974), for each manifest variable $u \in U$, K has a partition into disjoint nonempty sets D_{uw}, $w \in W_u$. The set Θ consists of $\boldsymbol{\pi} \in \Theta'$ for which

$$\pi_{\mathbf{k}'ug} = \pi_{\mathbf{k}ug} , \qquad \mathbf{k}, \mathbf{k}' \in D_{uw},\ w \in W_u,\ g \in L_u,\ u \in U .$$

Given (2.28) and (2.31),

$$\pi_{\mathbf{k}}^{(v+1)} = N^{-1}m_{\mathbf{k}}(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*)\,, \qquad \mathbf{k} \in K\,, \tag{4.13}$$

and

$$\textstyle\sum_{\mathbf{k}' \in D_{uw}} m_{\mathbf{k}'ug}(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*) = \pi_{\mathbf{k}ug}^{(v+1)} \sum_{\mathbf{k}' \in D_{uw}} m_{\mathbf{k}'}(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*)\,, \tag{4.14}$$
$$g \in L_u,\ \mathbf{k} \in D_{uw},\ w \in W_u,\ u \in U\,.$$

If $\Theta = \Theta'$, (4.14) reduces to

$$m_{\mathbf{k}ug}(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*) = \pi_{\mathbf{k}ug}^{(v+1)} m_k(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*)\,. \tag{4.15}$$

5. Solution of the likelihood equations by the Newton–Raphson and scoring algorithms. The functional iteration algorithm of the preceding section has the virtues of simplicity and stability, but it can be rather slow to converge and it provides no assistance in estimation of asymptotic covariance matrices. In contrast, the Newton–Raphson and scoring algorithms are generally more difficult to implement, but they provide rapid convergence when good initial estimates of $\hat{\boldsymbol{\pi}}$ are available and they facilitate estimation of $\Sigma(\boldsymbol{\pi} \mid \mathbf{e}^*)$. These latter properties may make the Newton–Raphson and scoring algorithms preferable, at least when asymptotic covariances are of interest.

To apply the Newton–Raphson algorithm, note that for $\mathbf{z} \in \Omega'(\hat{\boldsymbol{\pi}})$,

$$(\mathbf{z},\ \Pi^-(\hat{\boldsymbol{\pi}})\mathbf{m}(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*)) = 0\,. \tag{5.1}$$

The Newton–Raphson algorithm for solution of (5.1) uses an initial estimate $\boldsymbol{\pi}^{(0)} \in \Theta$ to generate a sequence $\{\pi^{(v)}\} \subset \Theta$ such that for $\mathbf{z} \in \Omega'(\boldsymbol{\pi}^{(v)})$,

$$(\mathbf{z},\ E(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*)(\boldsymbol{\pi}^{(v+1)} - \boldsymbol{\pi}^{(v)})) = (\mathbf{z},\ \Pi^-(\boldsymbol{\pi}^{(v)})\mathbf{m}(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*))\,. \tag{5.2}$$

If one proceeds as in the proof of Theorem 3, one finds that

$$\begin{aligned} \boldsymbol{\pi}^{(v+1)} &= \boldsymbol{\pi}^{(v)} + P(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*)[E(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*)]^-[P(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*)]^A\Pi^-(\boldsymbol{\pi}^{(v)})\mathbf{m}(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*) \\ &= \boldsymbol{\pi}^{(v)} + \Sigma(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*)\Pi^-(\boldsymbol{\pi}^{(v)})\mathbf{m}(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*) \end{aligned} \tag{5.3}$$

provided $(\mathbf{z}, E(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*)\mathbf{z}) > 0$ for all $\mathbf{z} \in \Omega'(\boldsymbol{\pi}^{(v)})$, $\mathbf{z} \neq \mathbf{0}$.

If $(\mathbf{z}, E(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*)\mathbf{z}) > 0$ for $\mathbf{z} \in \Omega'(\hat{\boldsymbol{\pi}})$, $\mathbf{z} \neq \mathbf{0}$, if $\pi_h^{(0)} = 0$ if $\hat{\pi}_h = 0$ and if $\boldsymbol{\pi}^{(0)}$ is sufficiently close to $\hat{\boldsymbol{\pi}}$, then $\boldsymbol{\pi}^{(v)} \to \hat{\boldsymbol{\pi}}$ and $\{\boldsymbol{\pi}^{(v)}\}$ has the quadratic convergence property $||\boldsymbol{\pi}^{(v+1)} - \hat{\boldsymbol{\pi}}|| \leqq c||\boldsymbol{\pi}^{(v)} - \hat{\boldsymbol{\pi}}||^2$ for some $c \geqq 0$ (see Ostrowski (1966, pages 183–194)). The large-sample analogue is that if the conditions of Theorem 3 hold, if $\boldsymbol{\pi}^{(0)} = T(\mathbf{n}^*)$, where $T(\mathbf{n}^*)$ converges to $\boldsymbol{\pi}$ in probability and $P\{T_h(\mathbf{n}^*) = 0\} \to 1$ if $\pi_h = 0$, then the probability approaches 1 that $\boldsymbol{\pi}^{(v)} \to \hat{\boldsymbol{\pi}}(\mathbf{n}^*)$. If in addition, $N^{\frac{1}{2}}[T(\mathbf{n}^*) - \boldsymbol{\pi}]$ converges to $\mathbf{0}$ in probability, then $N^{\frac{1}{2}}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^{(1)})$ converges to $\mathbf{0}$ in probability, so that $\boldsymbol{\pi}^{(1)}$ and $\hat{\boldsymbol{\pi}}$ have the same asymptotic distribution. If Z satisfies condition (B) in Section 3.1 and $\boldsymbol{\pi}^{(0)} = Z(N^{-1}\mathbf{n}^*)$, then convergence is quite rapid, for $\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^{(1)}$ is then of order $1/N$.

If sample sizes are small, then convergence is not assured. Stability can be produced by calculation of $\bar{\boldsymbol{\pi}}(f(\boldsymbol{\pi}^{(v)}))$. One sets $\boldsymbol{\pi}^{(v+1)} = \bar{\boldsymbol{\pi}}(f(\boldsymbol{\pi}^{(v)}))$ if use of (5.3) results in a $\boldsymbol{\pi}^{(v+1)}$ not in Θ_+ or if use of (5.3) results in a smaller value of $l(\mathbf{n}^*,$

$\boldsymbol{\pi}^{(v+1)}$). In this way, the stability of functional iteration and the speed of the Newton–Raphson algorithm may be combined.

An incidental benefit of the Newton–Raphson algorithm is that $\Sigma(\hat{\boldsymbol{\pi}} \mid \mathbf{n}^*)$ is an estimate of $\Sigma(\boldsymbol{\pi} \mid \mathbf{e}^*)$. This property may also be used in the scoring algorithm in which $\Sigma(\boldsymbol{\pi}^{(v)} \mid \mathbf{n}^*)$ in (5.3) is replaced by $\Sigma(\boldsymbol{\pi}^{(v)} \mid \{N_k p_j^*(\boldsymbol{\pi}^{(v)}) : j \in A_k,\ 1 \leqq k \leqq K\})$. If condition (B) of Section 3.1 holds, other conditions required in Theorem 3 are satisfied, and $\boldsymbol{\pi}^{(0)} = Z(N^{-1}\mathbf{n}^*)$, then it remains true in the scoring algorithm that $\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^{(1)}$ is of order $1/N$. Nonetheless, the scoring algorithm lacks the quadratic convergence property. The only reason for use of scoring is that in some problems, it results in simplifications in the computations for the Newton–Raphson algorithm.

Both the scoring and Newton–Raphson algorithms may be used with the examples considered in this paper. The programming labor is somewhat greater than in the case of functional iteration, but a substantial saving in computer time can be achieved with these algorithms.

REFERENCES

[1] Anderson, T. W. (1959). Some scaling models and estimation procedures in the latent class model. In *Probability and Statistics; the Harold Cramér Volume* (H. Grenander, ed.). Wiley, New York.

[2] Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about Markov chains. *Ann. Math. Statist.* **28** 89–110.

[3] 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.

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

[5] Chen, T. and Fienberg, S. E. (1974). Two-dimensional contingency tables with both completely and partially cross-classified data. *Biometrics* **30** 629–642.

[6] Cohen, J. E. (1971). Estimation and interaction in a censored $2 \times 2 \times 2$ contingency table. *Biometrics* **27** 379–386.

[7] Elandt-Johnson, R. C. (1971). *Probability Models and Statistical Methods in Genetics.* Wiley, New York.

[8] Fiacco, A. V. and McCormick, G. P. (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* Wiley, New York.

[9] Fienberg, S. E. and Larntz, K. (1971). Some models for individual-group comparisons and group behavior. *Psychometrika* **36** 349–367.

[10] Goodman, L. A. (1970). The multivariate analysis of qualitative data: interactions among multiple classifications. *J. Amer. Statist. Assoc.* **65** 226–256.

[11] Goodman, L. A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable: Part I—a modified latent structure approach. *Amer. J. Sociol.* **79** 1179–1259.

[12] Haberman, S. J. (1974). Log-linear models for frequency tables derived by indirect observation: maximum likelihood equations. *Ann. Statist.* **2** 911–924.

[13] Haberman, S. J. (1976). Iterative scaling procedures for log-linear models for frequency tables derived by indirect observation. *1975 Statistical Computing Section, Proc. Amer. Stat. Assoc.* 45–50.

[14] Halmos, P. (1958). *Finite-dimensional Vector Spaces.* Van Nostrand, Princeton.

[15] 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.

[16] LAZARSFELD, P. F. and HENRY, N. W. (1968). *Latent Structure Analysis.* Houghton Mifflin, Boston.

[17] LOOMIS, L. and STERNBERG, S. (1968). *Advanced Calculus.* Addison-Wesley, Reading, Mass.

[18] OSTROWSKI, A. M. (1967). *Solution of Equations and Systems of Equations.* Academic Press, New York.

[19] RAO, C. R. (1973). *Linear Statistical Inference and Its Applications.* Wiley, New York.

[20] RAO, C. R. and MITRA, S. K. (1971). *Generalized Inverse of Matrices and Its Applications.* Wiley, New York.

[21] SUNDBERG, R. (1971). Maximum likelihood theory and applications for distributions generated when observing a function of an exponential family variable. Dissertation, Institute of Mathematical Statistics, Stockholm.

[22] SUNDBERG, R. (1974). Maximum likelihood theory for incomplete data from an exponential family. *Scand. J. Statist.* **1** 49–58.

[23] ZANGWILL, W. (1969). *Nonlinear Programming: A Unified Approach.* Prentice-Hall, Englewood Cliffs, N.J.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 UNIVERSITY AVENUE
CHICAGO, ILLINOIS 60637