The Annals of Statistics
1973 Vol. 1, No. 4, 617–632

LOG-LINEAR MODELS FOR FREQUENCY DATA: SUFFICIENT STATISTICS AND LIKELIHOOD EQUATIONS[1]

BY SHELBY J. HABERMAN

University of Chicago

A general model is proposed for analysis of frequency tables. This model includes conventional log-linear models for complete and incomplete factorial tables and logit models for quantal response analysis. By use of coordinate-free methods of linear algebra and differential calculus, complete minimal sufficient statistics and likelihood equations for the maximum likelihood estimate are derived. The maximum likelihood estimate is shown to be unique if it exists, and necessary and sufficient conditions are given for its existence.

1. Introduction. Log-linear models for contingency tables have received considerable attention in recent years; however, with a few exceptions, discussion has been confined to models corresponding to linear models used in the analysis of variance. The log-linear models considered have not exploited ordering of categories or the existence of covariates, and necessary and sufficient conditions for existence of maximum likelihood estimates have not been given.

In this paper, models are considered which may be described in terms of linear manifolds. These models include the hierarchical log-linear models for factorial tables discussed by Bishop (1969), Fienberg (1970, 1972), and Goodman (1968, 1970), among others, together with the logit models of Finney (1952) and Dyke and Patterson (1952). The treatment in terms of linear manifolds permits development of a unified theory which allows examination of nonhierarchical log-linear models, models for ordered classifications, and multinomial response models.

The proposed models may be employed to analyze tables which result from Poisson or multinomial sampling. In Section 2, the model is defined and complete minimal sufficient statistics are found.

Maximum likelihood estimation is investigated in Section 3. No matter what sampling method is employed, maximum likelihood estimates are shown to be unique whenever they exist, and necessary and sufficient conditions are given for existence of these estimates. Maximum likelihood equations are given for the two sampling methods and estimates for Poisson and multinomial sampling are shown to coincide.

2. Basic properties of log-linear models. A log-linear model is used to describe

Received February 1972; revised December 1972.

[1] Research for this paper was supported in part by National Science Foundation Grant NSF 32037x. This paper is based on results in a dissertation written while the author was a John and Fannie Hertz Foundation Fellow at the University of Chicago.

AMS 1970 *subject classifications*. Primary 62B99, 62F10.

Key words and phrases. Log-linear models, contingency tables, logit models, maximum likelihood estimation, sufficient statistics.

a frequency table $\mathbf{n} = \{n_i : i \in I\} = \{n_i\}$ indexed by a finite nonempty set I containing q elements. This table is an element of the q-dimensional vector space R^I of real q-tuples $\mathbf{x} = \{x_i : i \in I\}$ with inner produce defined for $\mathbf{x}$ and $\mathbf{y}$ in R^I by

$$(\mathbf{x}, \mathbf{y}) = \textstyle\sum_{i \in I} x_i y_i \,. \tag{2.1}$$

In a log-linear model, $\mathbf{n}$ is assumed to have a mean $\mathbf{m} = \{m_i\}$ such that $m_i > 0$ for $i \in I$ and $\boldsymbol{\mu} = \{\log m_i\} \in \mathscr{M}$, where $\mathscr{M}$ is a p-dimensional linear manifold contained in R^I and $0 < p \leqq q$. Given this definition, the following are examples of log-linear models.

EXAMPLE 2.1. Consider an $r \times c$ contingency table with cell probabilities $p_{ij} > 0$, $1 \leqq i \leqq r$, $1 \leqq j \leqq c$, derived from a single multinomial sample of size N. If $\bar{r}$ is the set of integers from 1 to r and $\bar{c}$ is the set of integers from 1 to c, then $I = \bar{r} \times \bar{c}$, $\mathbf{n} = \{n_{ij} : (i, j) \in \bar{r} \times \bar{c}\}$, and the mean $\mathbf{m} = \{Np_{ij}\}$. If the variables represented by the rows and columns of the table are independent, then

$$p_{ij} = p_{i+} p_{+j} \,, \tag{2.2}$$

where the summation notation

$$p_{i+} = \textstyle\sum_{j=1}^{c} p_{ij} \tag{2.3}$$

and

$$p_{+j} = \textstyle\sum_{i=1}^{r} p_{ij} \tag{2.4}$$

is employed. As Bishop and Fienberg (1969) show, (2.2) is equivalent to the condition that $\boldsymbol{\mu} = \{\log m_{ij}\}$ be expressible in the form

$$\mu_{ij} = \alpha + \beta_i + \gamma_j \,, \tag{2.5}$$

where

$$\textstyle\sum_{i=1}^{r} \beta_i = \sum_{j=1}^{c} \gamma_j = 0 \,. \tag{2.6}$$

The set $\mathscr{M}$ of $\boldsymbol{\mu}$ which satisfy (2.5) and (2.6) for some α, $\{\beta_i\}$, and $\{\gamma_j\}$ is a linear manifold with dimension $r + c - 1$.

EXAMPLE 2.2. In a quantal response experiment, $N_j > 0$ subjects receive a log dosage t_j of a drug, where $1 \leqq j \leqq r$. Two responses 1 and 2 are possible. Of the N_j subjects, n_{j1} have response 1 and n_{j2} have response 2. If the probability that a subject given log dosage t has response 1 is $1/[1 + \exp -(\alpha + \beta t)]$, then a logit model is used for the data (see Finney (1952)). In this example, $I = \bar{r} \times \bar{2}$ and $\mathbf{n} = \{n_{jk} : (j, k) \in \bar{r} \times \bar{2}\}$. The assumption that the probability of response 1 given log dosage t_j is $1/[1 + \exp -(\alpha + \beta t)]$ is equivalent to the condition

$$\mu_{j1} - \mu_{j2} = \alpha + \beta t_j \,. \tag{2.7}$$

If $r \geqq 2$ and $t_j \neq t_{j'}$ if $j \neq j'$, then the set $\mathscr{M}$ of $\boldsymbol{\mu}$ such that (2.7) holds is a linear manifold of dimension $r + 2$.

EXAMPLE 2.3. A number of models have been used in the analysis of complete

$r \times c \times d$ tables with $I = \bar{r} \times \bar{c} \times \bar{d}$ and $\mathbf{n} = \{n_{ijk}\}$. In one model, the hypothesis of no three-factor interaction, $\boldsymbol{\mu}$ is assumed to satisfy the equation

$$(2.8) \qquad \mu_{ijk} = u + u_i{}^A + u_j{}^B + u_k{}^C + u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC}, \qquad \text{where}$$

$$(2.9) \qquad \sum_{i=1}^r u_i{}^A = \sum_{j=1}^c u_j{}^B = \sum_{k=1}^d u_k{}^C = \sum_{i=1}^r u_{ij}^{AB} = \sum_{j=1}^c u_{ij}^{AB} = \sum_{i=1}^r u_{ik}^{AC} = \sum_{k=1}^d u_{ik}^{AC} = \sum_{j=1}^c u_{jk}^{BC} = \sum_{k=1}^d u_{jk}^{BC} = 0 .$$

The set $\mathscr{M}$ of $\boldsymbol{\mu}$ which satisfy (2.8) and (2.9) for some u, $\{u_i{}^A\}$, $\{u_j{}^B\}$, $\{u_k{}^C\}$, $\{u_{ij}^{AB}\}$, $\{u_{ik}^{AC}\}$, and $\{u_{jk}^{BC}\}$ is a linear manifold of dimension $rc + rd + cd - r - c - d + 1$. This log-linear model has been examined by numerous authors. Goodman (1970) provides a thorough discussion of this model, as well as other related models.

Other examples may be constructed. The important point is that $\mathscr{M}$ is an arbitrary linear manifold; therefore, any linear model appropriate for linear regression or analysis of variance corresponds to a log-linear model.

No specification has been made yet concerning the underlying distribution of $\mathbf{n}$. In this paper, the principal probability models considered are the Poisson and multinomial models. A generalization of these models which is of some interest is the conditional Poisson model, which is discussed by Haberman (1970 and 1972).

2.1. *The Poisson model.* In the Poisson model, the elements of $\mathbf{n}$ are independent Poisson random variables with $E(n_i) = m_i$ for every $i \in I$ and $m_i > 0$ for each $i \in I$. If $\mathbf{m}(\boldsymbol{\mu}) = \{\exp \mu_i\}$ then the log likelihood may be written as

$$(2.10) \qquad l(\mathbf{n}, \boldsymbol{\mu}) = \sum_{i \in I} (n_i \log m_i(\boldsymbol{\mu}) - m_i(\boldsymbol{\mu}) - \log n_i!) = (\mathbf{n}, \boldsymbol{\mu}) - \sum_{i \in I} e^{\mu_i} - \sum_{i \in I} \log n_i!$$

In this equation, $\mathbf{n}$ is regarded as fixed and $l(\mathbf{n}, \boldsymbol{\mu})$ is a function defined for $\boldsymbol{\mu} \in \mathscr{M}$.

Let $P_{\mathscr{M}}$ be the orthogonal projection from R^I to $\mathscr{M}$. Since $\boldsymbol{\mu} \in \mathscr{M}$ and $P_{\mathscr{M}}$ is a symmetric operator,

$$(2.11) \qquad l(\mathbf{n}, \boldsymbol{\mu}) = (\mathbf{n}, P_{\mathscr{M}}\boldsymbol{\mu}) - \sum_{i \in I} e^{\mu_i} - \sum_{i \in I} \log n_i! = (P_{\mathscr{M}}\mathbf{n}, \boldsymbol{\mu}) - \sum_{i \in I} e^{\mu_i} - \sum_{i \in I} \log n_i!$$

Therefore, the family of Poisson models such that $\boldsymbol{\mu} \in \mathscr{M}$ is an exponential family. Since $\mathscr{M}$ and R^p are isomorphic (see Halmos (1958)), $P_{\mathscr{M}}n$ is a complete minimal sufficient statistic for $\boldsymbol{\mu}$. In addition, any nonsingular linear transformation of $P_{\mathscr{M}}\mathbf{n}$ is a complete minimal sufficient statistic. For example, if $\{\boldsymbol{\mu}^{(j)} : j \in \bar{s}\}$ spans $\mathscr{M}$, then $\{(\boldsymbol{\mu}^{(j)}, \mathbf{n}) : j \in \bar{s}\}$ is a complete minimal sufficient statistic.

EXAMPLE 2.4. Consider the hypothesis of no three-factor interaction of Example 2.3. It is readily shown that $\mathscr{M}$ is spanned by the vectors $\{\mathbf{x}^{(i,j)}\}$, $\{\mathbf{y}^{(i,k)}\}$, and $\{\mathbf{z}^{(j,k)}\}$, where

$$(2.12) \qquad x_{ijk}^{(i',j')} = 1 \quad \text{if} \quad i = i', \quad j = j', \qquad = 0 \quad \text{otherwise};$$

$$(2.13) \qquad y_{ijk}^{(i',j')} = 1 \quad \text{if} \quad i = i', \quad k = k', \qquad = 0 \quad \text{otherwise},$$

and

$$z_{ijk}^{(j',k')} = 1 \quad \text{if } j = j', \quad k = k', \qquad = 0 \quad \text{otherwise}. \tag{2.14}$$

The inner product

$$(\mathbf{n}, \mathbf{x}^{(i,j)}) = \sum_{k=1}^{r} n_{ijk} = n_{ij+}. \tag{2.15}$$

Similarly,

$$(\mathbf{n}, \mathbf{y}^{(i,k)}) = n_{i+k} \tag{2.16}$$

and

$$(\mathbf{n}, \mathbf{z}^{(j,k)}) = n_{+jk}. \tag{2.17}$$

Thus the marginal totals $\{n_{ij+}\}$, $\{n_{i+k}\}$, and $\{n_{+jk}\}$ form a complete minimal sufficient statistic under the Poisson model, as observed by Birch (1963) and Bishop (1969) and implicit in Darroch (1962).

2.2. *The multinomial model.* In the multinomial model, $I = \bigcup_{k \in \bar{s}} I_k$, where the I_k, $k \in \bar{s}$, are disjoint. For each $k \in \bar{s}$, $\{n_i : i \in I_k\}$ has a multinomial distribution with mean $\{m_i : i \in I_k\}$. The s collections of frequencies are independently distributed, and for each $k \in \bar{s}$, it is assumed that

$$\boldsymbol{\nu}^{(k)} = \{\chi_{I_k}(i) : i \in I\} \in \mathscr{M}, \tag{2.18}$$

where χ_{I_k} is the characteristic function of I_k.

Complete sufficient statistics may be found by considering a direct sum decomposition of $\mathscr{M}$ into $\mathscr{N}$ and $\mathscr{M} - \mathscr{N}$, where $\mathscr{N}$ is the manifold generated by $\{\boldsymbol{\nu}^{(k)} : k \in \bar{s}\}$ and $\mathscr{M} - \mathscr{N}$ is the orthogonal complement of $\mathscr{N}$ relative to $\mathscr{M}$ (see Kruskal (1968)). Suppose that the sample size for $\{n : i \in I_k\}$ is N_k. If $\boldsymbol{\mu} \in \mathscr{M}$ and $\{m_i(\boldsymbol{\mu}) : i \in I_k\}$ is consistent with N_k, then

$$(\mathbf{m}(\boldsymbol{\mu}), \boldsymbol{\nu}^{(k)}) = \sum_{i \in I_k} m_i(\boldsymbol{\mu}) = \sum_{i \in I_k} n_i = N_k. \tag{2.19}$$

For $\mathbf{x}$ and $\mathbf{y}$ in R^I, define $\mathbf{x} \cdot \mathbf{y}$ by $\{x_i y_i\}$. Then

$$\mathbf{m}(\boldsymbol{\mu}) = \mathbf{m}(P_{\mathscr{N}} \boldsymbol{\mu}) \cdot \mathbf{m}(P_{\mathscr{M}-\mathscr{N}} \boldsymbol{\mu}). \tag{2.20}$$

For some $\{c_k : k \in \bar{s}\}$, where c_k is a constant for $k \in \bar{s}$, expression (2.18) and the fact that each i is an element of exactly one I_k imply that

$$P_{\mathscr{N}} \boldsymbol{\mu} = \sum_{k=1}^{s} c_k \boldsymbol{\nu}^{(k)} = \{\sum_{k=1}^{s} c_k \chi_{I_k}(i) : i \in I\} = \{c_k : i \in I_k \text{ and } k \in \bar{s}\}. \tag{2.21}$$

Thus

$$\mathbf{m}(P_{\mathscr{N}} \boldsymbol{\mu}) = \{e^{c_k} : i \in I_k \text{ and } k \in \bar{s}\} \tag{2.22}$$

and

$$(\mathbf{m}(\boldsymbol{\mu}), \boldsymbol{\nu}^{(k)}) = (\mathbf{m}(P_{\mathscr{N}}\boldsymbol{\mu}) \cdot \mathbf{m}(P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu}), \boldsymbol{\nu}^{(k)}) \tag{2.23}$$
$$= \sum_{i \in I_k} m_i(P_{\mathscr{N}}\boldsymbol{\mu}) m_i(P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu})$$
$$= \sum_{i \in I_k} e^{c_k} m_i(P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu})$$
$$= e^{c_k}(\mathbf{m}(P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu}), \boldsymbol{\nu}^{(k)}) .$$

If $i \in I_k$, then $m_i(P_{\mathscr{N}}\boldsymbol{\mu}) = e^{c_k}$ and (2.19) and (2.23) imply that

$$m_i(P_{\mathscr{N}}\boldsymbol{\mu}) = \frac{N_k}{(\mathbf{m}(P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu}), \boldsymbol{\nu}^{(k)})} . \tag{2.24}$$

Thus $P_{\mathscr{N}}\boldsymbol{\mu}$ is a function of $P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu}$. Since $\boldsymbol{\mu}$ is $P_{\mathscr{N}}\boldsymbol{\mu} + P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu}$, $\boldsymbol{\mu}$ is a function of $P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu}$. A complete minimal sufficient statistic for $P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu}$ is consequently a complete minimal sufficient statistic for $\boldsymbol{\mu}$.

The log likelihood function is

$$l^{(m)}(\mathbf{n}, \boldsymbol{\mu})$$
$$= \sum_{k=1}^{s} \left[\sum_{i \in I_k} n_i \log \frac{m_i(\boldsymbol{\mu})}{(\mathbf{m}(\boldsymbol{\mu}), \boldsymbol{\nu}^{(k)})} + \log N_k! - \sum_{i \in I_k} \log n_i! \right]$$
$$= \sum_{k=1}^{s} \left[\sum_{i \in I_k} n_i \log \frac{m_i(P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu})}{(\mathbf{m}(P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu}), \boldsymbol{\nu}^{(k)})} + \log N_k! - \sum_{i \in I_k} \log n_i! \right] \tag{2.25}$$
$$= (P_{\mathscr{M}-\mathscr{N}}\mathbf{n}, P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu}) - \sum_{k=1}^{s} N_k \log (\mathbf{m}(P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu}), \boldsymbol{\nu}^{(k)})$$
$$+ \sum_{k=1}^{s} \log N_k! - \sum_{i \in I} \log n_i!$$

If $p = \dim \mathscr{M} > s$, then $\mathscr{M} - \mathscr{N}$ is isomorphic to R^{p-s}. The family of distributions such that $\boldsymbol{\mu} \in \mathscr{M}$ and for each $k \in \bar{s}$, $(\mathbf{m}(\boldsymbol{\mu}), \boldsymbol{\nu}^{(k)}) = N_k$, is then an exponential family with $P_{\mathscr{M}-\mathscr{N}}\mathbf{n}$ as a complete minimal sufficient statistic for $P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu}$. If $p = s$, then the family contains only one distribution. As in the Poisson model, alternate complete sufficient statistics may be obtained by use of nonsingular linear transformations. In particular, $P_{\mathscr{M}}\mathbf{n}$ is a complete minimal sufficient statistic.

EXAMPLE 2.5. In Example 2.1, a single multinomial sample is present, so that $s = 1$ and $I_1 = I = \bar{r} \times \bar{c}$. The vector $\boldsymbol{\nu}^{(1)}$ is the unit vector $\mathbf{e} = \{1\}$, which is an element of the manifold $\mathscr{M}$ defined by (2.5) and (2.6). Thus the model proposed in Example 2.1 is a multinomial model. Since $P_{\mathscr{M}}\mathbf{n}$ is a complete minimal sufficient statistic, an argument similar to that in Example 2.4 shows that the marginal totals $\{n_{i+}\}$ and $\{n_{+j}\}$ form a complete minimal sufficient statistic.

EXAMPLE 2.6. In Example 2.2, if the N_j are fixed, then multinomial sampling is employed with $s = r$ and $I_k = \{(k, 1), (k, 2)\}$. The vector $\boldsymbol{\nu}^{(k)}$ satisfies

$$\nu_{j1}^{(k)} - \nu_{j2}^{(k)} = 0 = 0 + 0t_j , \tag{2.26}$$

so $\boldsymbol{\nu}^{(k)} \in \mathscr{M}$ for each $k \in \bar{s}$. As shown in Haberman (1972),

$$\mathscr{M} - \mathscr{N} = \operatorname{span}\{\mathbf{x}, \mathbf{y}\} , \tag{2.27}$$

where $x_{j1} = 1$ and $x_{j2} = -1$ for $j \in \bar{r}$ and $y_{j1} = t_j$ and $y_{j2} = -t_j$ for $j \in \bar{r}$. Thus

$\{(\mathbf{x}, \mathbf{n}), (\mathbf{y}, \mathbf{n})\}$ is a complete minimal sufficient statistic for $\boldsymbol{\mu}$. Since $n_{j2} = N_j - n_{j1}$,

$$(2.28)\qquad (\mathbf{x}, \mathbf{n}) = \sum_{j=1}^{r} (n_{j1} - n_{j2}) = 2n_{+1} - N_{+}$$

and

$$(2.29)\qquad (\mathbf{y}, \mathbf{n}) = \sum_{j=1}^{r} t_j(n_{j1} - n_{j2}) = 2\sum_{j=1}^{r} n_{j1} t_j - \sum_{j=1}^{r} N_j t_j .$$

Consequently $(n_{+1}, \sum_{j=1}^{r} n_{j1} t_j)$ is a complete minimal sufficient statistic.

The complete minimal sufficient statistics found in this section can be applied in some cases to construction of minimum variance unbiased estimates and exact confidence intervals and hypothesis tests (see Lehmann (1959) and Haberman (1972)); however, the principal application of these statistics is in maximum likelihood estimation, the subject of the next section.

3. Maximum likelihood estimation. It is convenient to begin consideration of maximum likelihood estimation with an examination of the Poisson model. Results for the multinomial model follow directly. In this section, existence and uniqueness of estimates is investigated. This topic has been considered by Birch (1963) in connection with hierarchical models for complete factorial tables such as the model in Example 2.3. More recently, Fienberg (1970, 1972) has considered the problem in the case of incomplete multiway tables. Results in this section are more general and sharper than those previously derived.

3.1. *The Poisson model.* In the maximum likelihood estimation problem for the Poisson model, an element $\hat{\boldsymbol{\mu}}$ of $\mathscr{M}$ is sought such that

$$(3.1)\qquad l(\mathbf{n}, \hat{\boldsymbol{\mu}}) = \sup_{\boldsymbol{\mu} \in \mathscr{M}} l(\mathbf{n}, \boldsymbol{\mu}) .$$

If $\hat{\boldsymbol{\mu}}$ exists, then it is a maximum likelihood estimate (MLE) of $\boldsymbol{\mu}$ and $\hat{\mathbf{m}} = \{\exp \hat{\mu}_i\}$ is an MLE of $\mathbf{m}$.

In order to examine the properties of $\hat{\boldsymbol{\mu}}$, it is necessary to consider the first and second differentials of $l(\mathbf{n}, \boldsymbol{\mu})$. The first differential at $\boldsymbol{\mu}$ is a linear function $dl_{\boldsymbol{\mu}}(\mathbf{n}, \boldsymbol{\nu})$ defined for $\boldsymbol{\nu} \in \mathscr{M}$ such that

$$(3.2)\qquad l(\mathbf{n}, \boldsymbol{\mu} + \boldsymbol{\nu}) = l(\mathbf{n}, \boldsymbol{\mu}) + dl_{\boldsymbol{\mu}}(\mathbf{n}, \boldsymbol{\nu}) + o(\boldsymbol{\nu}) ,$$

where $o(\boldsymbol{\nu})/||\boldsymbol{\nu}|| \to 0$ as $||\boldsymbol{\nu}|| \to 0$. By elementary calculus,

$$(3.3)\qquad n(x + y) - e^{x+y} = (nx - e^x) + (ny - e^x y) + o(y) ,$$

where $(1/y)o(y) \to 0$ as $y \to 0$. Therefore,

$$(3.4)\qquad dl_{\boldsymbol{\mu}}(\mathbf{n}, \boldsymbol{\nu}) = (\boldsymbol{\nu}, \mathbf{n} - \mathbf{m}(\boldsymbol{\mu})) .$$

The second differential $d^2 l_{\boldsymbol{\mu}}(\mathbf{n}, \boldsymbol{\xi})(\boldsymbol{\nu})$ of $l(\mathbf{n}, \boldsymbol{\mu})$ at $\boldsymbol{\mu}$ is a linear function from $\mathscr{M}$ to the space $\mathscr{M}^*$ of linear functionals on $\mathscr{M}$. This differential satisfies

$$(3.5)\qquad dl_{\boldsymbol{\mu}+\boldsymbol{\xi}}(\mathbf{n}, \boldsymbol{\nu}) = dl_{\boldsymbol{\mu}}(\mathbf{n}, \boldsymbol{\nu}) + d^2 l_{\boldsymbol{\mu}}(\mathbf{n}, \boldsymbol{\xi})(\boldsymbol{\nu}) + o_{\boldsymbol{\xi}}(\boldsymbol{\nu}) ,$$

where

$$\lim_{||\boldsymbol{\xi}||\to 0} \frac{1}{||\boldsymbol{\xi}||} \sup_{||\boldsymbol{\nu}||=1} ||o_{\boldsymbol{\xi}}(\boldsymbol{\nu})|| \to 0 \,. \tag{3.6}$$

Since

$$\begin{aligned} dl_{\mu+\xi}(\mathbf{n}, \boldsymbol{\nu}) - dl_{\mu}(\mathbf{n}, \boldsymbol{\nu}) &= \textstyle\sum_{i\in I} \nu_i(e^{\mu_i} - e^{\mu_i+\xi_i}) \\ &= - \textstyle\sum_{i\in I} \nu_i \xi_i e^{\mu_i} + (\boldsymbol{\nu}, \mathbf{o}(\boldsymbol{\xi}))\,, \end{aligned} \tag{3.7}$$

where $(1/||\boldsymbol{\xi}||)||\mathbf{o}(\boldsymbol{\xi})|| \to 0$ as $||\boldsymbol{\xi}|| \to 0$, it follows that

$$d^2 l_{\mu}(\mathbf{n}, \boldsymbol{\xi})(\boldsymbol{\nu}) = - \textstyle\sum_{i\in I} \nu_i \xi_i e^{\mu_i} = - \sum_{i\in I} \nu_i \xi_i m_i(\boldsymbol{\mu})\,. \tag{3.8}$$

If $D(\boldsymbol{\mu})\{x_i\} = \{e^{\mu_i} x_i\}$, then one may write

$$d^2 l_{\mu}(\mathbf{n}, \boldsymbol{\xi})(\boldsymbol{\nu}) = -(\boldsymbol{\nu}, D(\boldsymbol{\mu})\boldsymbol{\xi})\,. \tag{3.9}$$

If $\boldsymbol{\nu} \neq 0$, $\boldsymbol{\nu} \in \mathscr{M}$, and $\boldsymbol{\mu} \in \mathscr{M}$, then

$$d^2 l_{\mu}(\mathbf{n}, \boldsymbol{\nu})(\boldsymbol{\nu}) = - \textstyle\sum_{i\in I} \nu_i^2 e^{\mu_i} < 0\,. \tag{3.10}$$

Thus $l(\mathbf{n}, \boldsymbol{\mu})$ is a strictly concave function of $\boldsymbol{\mu}$.

Given the results of the preceding paragraph, the following theorem can be proven:

THEOREM 3.1. *If an* MLE $\hat{\boldsymbol{\mu}}$ *exists, then it is unique and satisfies the equation*

$$P_{\mathscr{M}} \hat{\mathbf{m}} = P_{\mathscr{M}} \mathbf{n}\,. \tag{3.11}$$

Conversely, if for some $\hat{\boldsymbol{\mu}} \in \mathscr{M}$ *and* $\hat{\mathbf{m}} = \{e^{\hat{\mu}_i}\}$, (3.11) *is satisfied, then* $\hat{\boldsymbol{\mu}}$ *is the* MLE *of* $\boldsymbol{\mu}$.

PROOF. Since $l(\mathbf{n}, \boldsymbol{\mu})$ is strictly concave, at most one critical point exists, and this point must be a maximum. Therefore, only one MLE can exist. If the MLE $\hat{\boldsymbol{\mu}}$ exists, then for every $\boldsymbol{\nu} \in \mathscr{M}$,

$$dl_{\hat{\mu}}(\mathbf{n}, \boldsymbol{\nu}) = (\boldsymbol{\nu}, \mathbf{n} - \hat{\mathbf{m}}) = 0\,, \tag{3.12}$$

$\mathbf{n} - \hat{\mathbf{m}} \in \mathscr{M}^{\perp} = \{\mathbf{x} \in R^I : (\mathbf{x}, \boldsymbol{\mu}) = 0\ \forall \boldsymbol{\mu} \in \mathscr{M}\}$, and equation (3.11) must hold. On the other hand, if $\hat{\mathbf{m}}$ satisfies equation (3.11), then

$$dl_{\hat{\mu}}(\mathbf{n}, \boldsymbol{\nu}) = (\boldsymbol{\nu}, \mathbf{n} - \hat{\mathbf{m}}) = (P_{\mathscr{M}}\boldsymbol{\nu}, \mathbf{n} - \hat{\mathbf{m}}) = (\boldsymbol{\nu}, P_{\mathscr{M}}\mathbf{n} - P_{\mathscr{M}}\hat{\mathbf{m}}) = 0 \tag{3.13}$$

for every $\boldsymbol{\nu} \in \mathscr{M}$. Thus a critical point exists at $\hat{\boldsymbol{\mu}}$. □

The likelihood equation (3.11) requires that $\hat{\mathbf{m}}$ fit the sufficient statistic $P_{\mathscr{M}}\mathbf{n}$. If $\{\boldsymbol{\mu}^{(j)} : j \in \bar{s}\}$ spans $\mathscr{M}$, then equation (3.11) is equivalent to the condition

$$(\hat{\mathbf{m}}, \boldsymbol{\mu}^{(j)}) = (\mathbf{n}, \boldsymbol{\mu}^{(j)}) \tag{3.14}$$

for every $j \in \bar{s}$. This equation is particularly suggestive if

$$\begin{aligned} \mu_i^{(j)} &= 1 \qquad \text{if} \quad i \in I_j\,, \\ &= 0 \qquad \text{if} \quad i \in I - I_j\,, \end{aligned} \tag{3.15}$$

in which case

$$\textstyle\sum_{i\in I_j} \hat{m}_i = \sum_{i\in I_j} n_i\,. \tag{3.16}$$

Thus certain marginal totals must be equal for $\hat{\mathbf{m}}$ and $\mathbf{n}$. This relationship is frequently used in the discussion of hierarchical models by Birch (1963), Fienberg (1970, 1972), and Goodman (1968, 1970), among others, although general proofs are not provided in these references.

So far, no conditions have been given for the existence of the maximum likelihood estimate. In order to rectify this situation, the following theorem is useful:

THEOREM 3.2. *A necessary and sufficient condition that the* MLE $\hat{\boldsymbol{\mu}}$ *of* $\boldsymbol{\mu}$ *exists is that there exist* $\boldsymbol{\delta} \in \mathscr{M}^{\perp}$ *such that* $n_i + \delta_i > 0$ *for every* $i \in I$.

PROOF. To prove necessity, assume that $\hat{\boldsymbol{\mu}}$ satisfies equation (3.11). Then $\hat{\mathbf{m}} - \mathbf{n} \in \mathscr{M}^{\perp}$. In addition, $\hat{\mathbf{m}} - \mathbf{n} + \mathbf{n} = \hat{\mathbf{m}}$, where $\hat{m}_i > 0$ for each $i \in I$. Thus $\boldsymbol{\delta} = \hat{\mathbf{m}} - \mathbf{n}$ has the desired properties.

To prove sufficiency, assume that there exists $\boldsymbol{\delta} \in \mathscr{M}^{\perp}$ such that $n_i + \delta_i > 0$ for each $i \in I$. Suppose that

$$(3.17) \qquad \hat{l}^{(p)}(\mathbf{n}, \boldsymbol{\mu}) = \textstyle\sum_{i \in I} (n_i \mu_i - e^{\mu_i}) = (\mathbf{n}, \boldsymbol{\mu}) - \sum_{i \in I} e^{\mu_i} .$$

Then $\hat{l}^{(p)}(\mathbf{n}, \boldsymbol{\mu})$ and $l(\mathbf{n}, \boldsymbol{\mu})$ differ only by a constant. Since $\boldsymbol{\delta} \in \mathscr{M}^{\perp}$, $(\mathbf{n}, \boldsymbol{\mu}) = (\mathbf{n} + \boldsymbol{\delta}, \boldsymbol{\mu})$ and

$$(3.18) \qquad \hat{l}^{(p)}(\mathbf{n}, \boldsymbol{\mu}) = \textstyle\sum_{i \in I} [(n_i + \delta_i)\mu_i - e^{\mu_i}] .$$

Each summand is bounded above. Therefore, if any summand is small enough, then $\hat{l}^{(p)}(\mathbf{n}, \boldsymbol{\mu}) < \hat{l}^{(p)}(\mathbf{n}, \mathbf{0})$. For any $i \in I$ $n_i + \delta_i > 0$. Thus as $|\mu_i| \to \infty$, $(n_i + \delta_i)\mu_i - e^{\mu_i} \to -\infty$. Suppose $A = \{\boldsymbol{\mu} \in \mathscr{M}: \hat{l}^{(p)}(\mathbf{n}, \boldsymbol{\mu}) \geqq \hat{l}^{(p)}(\mathbf{n}, \mathbf{0})\}$. Then A is bounded. Since $\hat{l}^{(p)}(\mathbf{n}, \boldsymbol{\mu})$ is continuous in $\boldsymbol{\mu}$, A is closed. Therefore, $\hat{l}^{(p)}(\mathbf{n}, \boldsymbol{\mu})$ has a finite maximum for some $\boldsymbol{\mu} \in A$. □

The following corollary follows immediately:

COROLLARY 3.1. *If* $n_i > 0$ *for every* $i \in I$, *then the* MLE $\hat{\boldsymbol{\mu}}$ *exists.*

PROOF. Use $\boldsymbol{\delta} = \mathbf{0}$ and apply Theorem 3.2. □

A related condition to that of Theorem 3.2 is often useful:

THEOREM 3.3. *A necessary and sufficient condition that the* MLE $\hat{\boldsymbol{\mu}}$ *exist is that there not exist* $\boldsymbol{\mu} \in \mathscr{M}$ *such that* $\boldsymbol{\mu} \neq \mathbf{0}$, $\mu_i \leqq 0$ *for every* $i \in I$, *and* $(\mathbf{n}, \boldsymbol{\mu}) = 0$.

PROOF. Suppose that the MLE of $\boldsymbol{\mu}$ exists. Then there exists $\boldsymbol{\delta} \in \mathscr{M}^{\perp}$ such that $n_i + \delta_i > 0$ for each $i \in I$. If $\boldsymbol{\mu} \in \mathscr{M}$, $(\mathbf{n}, \boldsymbol{\mu}) = 0$, $\mu_i \leqq 0$ for every $i \in I$, and $\boldsymbol{\mu} \neq \mathbf{0}$, then $(\mathbf{n} + \boldsymbol{\delta}, \boldsymbol{\mu}) < 0$. Since $(\mathbf{n}, \boldsymbol{\mu})$ and $(\mathbf{n} + \boldsymbol{\delta}, \boldsymbol{\mu})$ are equal, a contradiction results. Thus no such $\boldsymbol{\mu}$ exists.

On the other hand, suppose that the MLE does not exist. Then there does not exist $\boldsymbol{\delta} \in \mathscr{M}^{\perp}$ such that $n_i + \delta_i > 0$ for every $i \in I$. Let I_0 be the set of indices in I such that $n_i = 0$. Suppose $S = \{\mathbf{x} \in R^I : x_i > 0 \; \forall i \in I_0\}$. Then S and $\mathscr{M}^{\perp}$ are disjoint convex sets. By the separating hyperplane theorem (see Blackwell and Girshick (1954)), there exists $\boldsymbol{\mu} \in R^I$, $\boldsymbol{\mu} \neq \mathbf{0}$, such that if $\boldsymbol{\nu} \in \mathscr{M}^{\perp}$, $(\boldsymbol{\mu}, \boldsymbol{\nu}) \geqq 0$, and if $\boldsymbol{\nu} \in S$, $(\boldsymbol{\mu}, \boldsymbol{\nu}) \leqq 0$. Suppose $j \in I - I_0$. If $\boldsymbol{\nu} \in S$, then $\boldsymbol{\nu} + c\boldsymbol{\delta}^{(j)} \in S$ for any real-valued c, where $\boldsymbol{\delta}^{(j)} = \{\delta_{ij} : i \in I\}$ and δ_{ij} is the Kronecker delta. Thus

$(\boldsymbol{\mu}, \boldsymbol{\delta}^{(j)}) = \mu_j = 0$. Now suppose $j \in I_0$. Then if $\boldsymbol{\nu} \in S$, $\boldsymbol{\nu} + c\boldsymbol{\delta}^{(j)} \in S$ for all positive c. Thus $(\boldsymbol{\mu}, \boldsymbol{\delta}^{(j)}) = \mu_j \leqq 0$. It is now sufficient to show that $\boldsymbol{\mu} \in \mathscr{M}$. This result follows since if $P_{\mathscr{M}^\perp}$ is the orthogonal projection from R^I to $\mathscr{M}^\perp$, then

$$(3.19) \qquad (\boldsymbol{\mu}, -P_{\mathscr{M}^\perp}\boldsymbol{\mu}) = -\|P_{\mathscr{M}^\perp}\boldsymbol{\mu}\|^2 \geqq 0 .$$

Unless $P_{\mathscr{M}^\perp}\boldsymbol{\mu}$ is $\mathbf{0}$, there is a contradiction. □

The following corollaries may be proven:

COROLLARY 3.2. *Suppose M_1 and M_2 are linear manifolds such that $M_1 \subset M_2$. Suppose $\hat{\boldsymbol{\mu}}^{(i)}$ is the* MLE *for $\boldsymbol{\mu} \in \mathscr{M}_i$, $i \in \bar{2}$. If $\hat{\boldsymbol{\mu}}^{(2)}$ exists, then $\hat{\boldsymbol{\mu}}^{(1)}$ exists. If $\hat{\boldsymbol{\mu}}^{(1)}$ does not exist, then $\hat{\boldsymbol{\mu}}^{(2)}$ does not exist.*

PROOF. If $\hat{\boldsymbol{\mu}}^{(2)}$ exists, then there exists $\boldsymbol{\delta} \in \mathscr{M}_2^\perp$ such that $n_i + \delta_i > 0$ for each $i \in I$. Since $\mathscr{M}_2^\perp \subset M_1^\perp$, $\boldsymbol{\delta} \in \mathscr{M}_1^\perp$. Therefore, $\hat{\boldsymbol{\mu}}^{(1)}$ exists. The converse follows immediately. □

COROLLARY 3.3. *Suppose $I_1 = \{i \in I: n_i = 0\}$. Let $\rho: R^I \to R^{I_1}$ satisfy $\rho\{\mu_i: i \in I\} = \{\mu_i: i \in I_1\}$. Let $\rho(\mathscr{M}^\perp) = \{\rho(\boldsymbol{\mu}): \boldsymbol{\mu} \in \mathscr{M}^\perp\}$. If $\rho(\mathscr{M}^\perp) = R^{I_1}$, then the* MLE *of $\boldsymbol{\mu}$ exists.*

PROOF. Suppose that $(\mathbf{n}, \boldsymbol{\mu}) = 0$, $\boldsymbol{\mu} \in \mathscr{M}$, $\boldsymbol{\mu} \neq \mathbf{0}$, and $\mu_i \leqq 0$ for every $i \in I$. Then $\mu_i = 0$ for every $i \in I - I_1$. If $((\cdot, \cdot))$ denotes the standard inner product for R^{I_1}, then $((\rho(\boldsymbol{\mu}), \rho(\boldsymbol{\nu}))) = (\boldsymbol{\mu}, \boldsymbol{\nu}) = 0$ for $\boldsymbol{\nu} \in \mathscr{M}^\perp$. Thus $\rho(\boldsymbol{\mu})$ is orthogonal to all elements of $\rho(\mathscr{M}^\perp) = R^{I_1}$. Therefore, $\rho(\boldsymbol{\mu}) = \mathbf{0}$. Thus $\boldsymbol{\mu} = \mathbf{0}$, a contradiction. Hence the MLE exists. □

COROLLARY 3.4. *Suppose $I_0 = \{i \in I: n_i > 0\}$. Let $\pi: R^I \to R^{I_0}$ satisfy $\pi\{\mu_i: i \in I\} = \{\mu_i: i \in I_0\}$. Let $\pi(\mathscr{M}) = \{\pi(\mu): \mu \in \mathscr{M}\}$. Suppose $\pi(\mathscr{M})$ has dimension k and $I - I_0$ has h elements. If $p - k = h$, then the* MLE *of $\boldsymbol{\mu}$ does not exist. If $p - k = 0$, then the* MLE *exists.*

PROOF. The kernel of π has dimension $p - k$. This kernel is a submanifold of $\{\boldsymbol{\mu} \in R^I: \mu_i = 0 \ \forall i \in I_0\}$. This latter manifold has dimension h. If $p - k = h$, then the kernel is equal to $\{\boldsymbol{\mu} \in R^I: \mu_i = 0 \ \forall i \in I_0\}$. Hence a $\boldsymbol{\mu} \in \mathscr{M}$ exists such that $\mu_i \leqq 0$ for each $i \in I$, $(\mathbf{n}, \boldsymbol{\mu}) = 0$, and $\boldsymbol{\mu} \neq \mathbf{0}$. By Theorem 3.3, the MLE of $\boldsymbol{\mu}$ does not exist. If $p - k = 0$, then the kernel of π is $\mathbf{0}$. Since no $\boldsymbol{\mu} \in \mathscr{M}$ which is not equal to $\mathbf{0}$ exists such that $(\mathbf{n}, \boldsymbol{\mu}) = 0$ and $\mu_i \leqq 0$ for $i \neq I$, the MLE exists. □

The following examples illustrate use of these theorems and corollaries in terms of the $r \times c \times d$ table of Example 2.3.

EXAMPLE 3.1. Theorem 3.1 and the results of Example 2.4 imply that if $\mathscr{M}$ is the linear manifold corresponding to the hypothesis of no three-factor interaction and if $\hat{\mathbf{m}}$ exists, then

$$(3.20) \qquad (\hat{\mathbf{m}}, \mathbf{x}^{(i,j)}) = \hat{m}_{ij+} = n_{ij+} ,$$

$$(3.21) \qquad \hat{m}_{i+k} = n_{i+k} ,$$

and

$$\hat{m}_{+jk} = n_{+jk} \tag{3.22}$$

(see Birch (1963)). Since $-\mathbf{x}^{(i,j)}$, $-\mathbf{y}^{(i,k)}$, and $-\mathbf{z}^{(j,k)}$ are all in $\mathscr{M}$, are nonzero, and have no positive coordinates, Theorem 3.3 implies that $\hat{\mathbf{m}}$ can only exist if $n_{ij+} > 0$ for $i \in \bar{r}$ and $j \in \bar{c}$, $n_{i+k} > 0$ for $i \in \bar{r}$ and $j \in \bar{d}$, and $n_{+jk} > 0$ for $j \in \bar{c}$ and $k \in \bar{d}$. These conditions are not, however, sufficient to ensure that $\hat{\mathbf{m}}$ exists, as will be shown in Example 3.2.

EXAMPLE 3.2. Suppose that in the preceding example, $r = c = d = 2$. Then $\mathscr{M}^{\perp}$ is the span of $\boldsymbol{\mu}^*$, where

$$\begin{aligned} \mu^*_{ijk} &= -1 \quad \text{if} \quad i+j+k \quad \text{is even}, \\ &= 1 \quad \text{if} \quad i+j+k, \quad \text{is odd}. \end{aligned} \tag{3.22}$$

Using tabular notation, one may write

$$\boldsymbol{\mu}^* = \begin{array}{|c|c|}\hline 1 & -1 \\ \hline -1 & 1 \\ \hline \end{array} \quad \begin{array}{|c|c|}\hline -1 & 1 \\ \hline 1 & -1 \\ \hline \end{array}. \tag{3.24}$$

In this representation, the left block represents $k = 1$ and the right block represents $k = 2$. The first row in each block stands for $i = 1$, and the first column stands for $j = 1$. Suppose $J = \{(i, j, k) : i + j + k \text{ is odd}\}$. Let $K = \{(i, j, k) : i + j + k \text{ is even}\}$. Then a MLE of $\boldsymbol{\mu}$ exists if and only if either $I_1 \subset J$ or $I_1 \subset K$, where I_1 is defined as in Corollary 3.3. To verify this result, observe that if $I_1 \subset J$, then $n_{ijk} + \frac{1}{2}\mu^*_{ijk} > 0$ for each $(i, j, k) \in \bar{2} \times \bar{2} \times \bar{2}$. Thus the maximum likelihood estimate exists. A similar result holds if $I_1 \subset K$. On the other hand, if $(i, j, k) \in J$, $(i', j', k') \in K$, and $n_{ijk} = n_{i'j'k'} = 0$, then for any real c,

$$n_{ijk} + c\mu^*_{ijk} = c \quad \text{and} \tag{3.25}$$

$$n_{i'j'k'} + c\mu^*_{i'j'k'} = -c. \tag{3.26}$$

Thus there exists no $\boldsymbol{\delta} \in \mathscr{M}^{\perp}$ such that $n_{ijk} + \delta_{ijk} > 0$ for every $(i, j, k) \in \bar{2} \times \bar{2} \times \bar{2}$. Hence the MLE does not exist.

To illustrate this result, it is useful to consider several different values of $\mathbf{n}$, employing the tabular notation of (3.24). If

$$\mathbf{n} = \begin{array}{|c|c|}\hline 0 & 8 \\ \hline 9 & 13 \\ \hline \end{array} \quad \begin{array}{|c|c|}\hline 4 & 8 \\ \hline 6 & 10 \\ \hline \end{array}, \tag{3.27}$$

then $I_1 \subset J$. Thus the MLE exists. In fact, whenever there is only one zero cell, the MLE exists.

When there are two cells which are zero, then the MLE may or may not exist. If

$$\mathbf{n} = \begin{array}{|c|c|}\hline 0 & 0 \\ \hline 9 & 13 \\ \hline \end{array} \quad \begin{array}{|c|c|}\hline 4 & 8 \\ \hline 6 & 10 \\ \hline \end{array}, \tag{3.28}$$

the estimate does not exist since $(1, 1, 1) \in J$ and $(1, 2, 1) \in K$. The result may also be verified by observing that the observed marginal total $n_{1+1} = 0$ and $\boldsymbol{\mu}^+ \in \mathscr{M}$, where

$$\boldsymbol{\mu}^+ = \begin{array}{|c|c|}\hline -1 & -1 \\ \hline 0 & 0 \\ \hline \end{array} \quad \begin{array}{|c|c|}\hline 0 & 0 \\ \hline 0 & 0 \\ \hline \end{array} . \tag{3.29}$$

Another case in which the MLE does not exist occurs when

$$\mathbf{n} = \begin{array}{|c|c|}\hline 0 & 8 \\ \hline 9 & 13 \\ \hline \end{array} \quad \begin{array}{|c|c|}\hline 4 & 8 \\ \hline 6 & 0 \\ \hline \end{array} . \tag{3.30}$$

Although the marginal totals $\{n_{ij+}\}$, $\{n_{i+k}\}$ and $\{n_{+jk}\}$ are all positive, $(1, 1, 1) \in J$ and $(2, 2, 2) \in K$. Thus $n_{111} + c\mu_{111}$ and $n_{222} + c\mu_{222}$ are of opposite signs. However, if

$$\mathbf{n} = \begin{array}{|c|c|}\hline 0 & 8 \\ \hline 9 & 0 \\ \hline \end{array} \quad \begin{array}{|c|c|}\hline 4 & 8 \\ \hline 6 & 10 \\ \hline \end{array} , \tag{3.31}$$

then the MLE does exist since both $(1, 1, 1)$ and $(2, 2, 1)$ are in J. The result also follows directly from Theorem 3.2 since the elements of $n + \boldsymbol{\mu}^*$ are all positive.

In general, rules for existence of MLE's are more difficult to find than in the preceding examples. Nevertheless, Theorems 3.2 and 3.3 and Corollaries 3.1, 3.2, 3.3, and 3.4 are readily applied to problems in which a specific $\mathbf{n}$ and $\mathscr{M}$ are considered. Further applications are given in Haberman (1970 and 1972). The following example illustrates a possible procedure for a $3 \times 3 \times 3$ table.

EXAMPLE 3.3. In Example 3.1, suppose that $r = c = d = 3$ and assume that in (2.8), $u_{1j}^{AB} = -u_{3j}^{AB}$, $1 \leqq j \leqq 3$. The likelihood equations are then

$$\hat{m}_{i+k} = n_{i+k} , \tag{3.32}$$

$$\hat{m}_{+jk} = n_{+jk} , \tag{3.33}$$

$$\hat{m}_{1j+} - \hat{m}_{3j+} = n_{1j+} - n_{3j+} . \tag{3.34}$$

Suppose

$$\mathbf{n} = \begin{array}{|c|c|c|}\hline 0 & 5 & 0 \\ \hline 3 & 0 & 4 \\ \hline 2 & 0 & 1 \\ \hline \end{array} \quad \begin{array}{|c|c|c|}\hline 0 & 3 & 0 \\ \hline 4 & 1 & 6 \\ \hline 0 & 1 & 3 \\ \hline \end{array} \quad \begin{array}{|c|c|c|}\hline 0 & 1 & 2 \\ \hline 2 & 1 & 0 \\ \hline 3 & 0 & 3 \\ \hline \end{array} . \tag{3.35}$$

To show that $\hat{\boldsymbol{\mu}}$ exists, first note that $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$ are in $\mathscr{M}^{\perp}$, where

$$\boldsymbol{\mu}^{(1)} = \begin{array}{|c|c|c|}\hline 1 & -1 & 0 \\ \hline -2 & 2 & 0 \\ \hline 1 & -1 & 0 \\ \hline \end{array} \quad \begin{array}{|c|c|c|}\hline 1 & -1 & 0 \\ \hline -2 & 2 & 0 \\ \hline 1 & -1 & 0 \\ \hline \end{array} \quad \begin{array}{|c|c|c|}\hline 1 & -1 & 0 \\ \hline -2 & 2 & 0 \\ \hline 1 & -1 & 0 \\ \hline \end{array} \tag{3.36}$$

and

$$\boldsymbol{\mu}^{(2)} = \begin{array}{|c|c|c|}\hline 0 & 0 & 0 \\ \hline 0 & -1 & 1 \\ \hline 0 & 1 & -1 \\ \hline \end{array} \quad \begin{array}{|c|c|c|}\hline 0 & 0 & 0 \\ \hline 0 & 2 & -2 \\ \hline 0 & -2 & 2 \\ \hline \end{array} \quad \begin{array}{|c|c|c|}\hline 0 & 0 & 0 \\ \hline 0 & -1 & 1 \\ \hline 0 & 1 & -1 \\ \hline \end{array} . \tag{3.37}$$

If $\delta > 0$ is sufficiently small, then $n_{i_1i_2i_3} + \delta(\mu^{(1)}_{i_1i_2i_3} + \mu^{(2)}_{i_1i_2i_3}) \geqq 0$ for $1 \leqq i_j \leqq 3$, $1 \leqq j \leqq 3$, and the inequality is strict except for the indices (1, 3, 1), (1, 3, 2), (3, 2, 1), and (3, 2, 3). Thus $\mathbf{n} + \delta(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)})$ has 6 fewer elements equal to 0 than $\mathbf{n}$ has. To eliminate the remaining four zeroes, define $\boldsymbol{\mu}^{(3)} \in \mathscr{M}^\perp$ by

(3.38) $\boldsymbol{\mu}^{(3)} =$

−2	1	1		−2	1	1		−2	1	1
4	−2	−2		4	−2	−2		4	−2	−2
−2	1	1		−2	1	1		−2	1	1

If $\gamma > 0$ is sufficiently small, then all elements of

$$\mathbf{n} + \delta(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)}) + \gamma\boldsymbol{\mu}^{(3)}$$

are positive. Since

$$\delta(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)}) + \gamma\boldsymbol{\mu}^{(3)} \in \mathscr{M}^\perp ,$$

the MLE exists.

3.2. *The multinomial model.* The multinomial model is closely related to the Poisson model. In the maximum likelihood estimation problem for the multinomial model, an element $\hat{\boldsymbol{\mu}}^{(m)}$ of $\tilde{\mathscr{M}}$ is sought such that

$$l^{(m)}(\mathbf{n}, \hat{\boldsymbol{\mu}}^{(m)}) = \sup_{\boldsymbol{\mu} \in \tilde{\mathscr{M}}} l^{(m)}(\mathbf{n}, \boldsymbol{\mu}) , \tag{3.39}$$

where $\tilde{\mathscr{M}} = \{\boldsymbol{\mu} \in \mathscr{M}: \sum_{i \in I_k} e^{\mu_i} = \sum_{i \in I_k} n_i \;\forall\, k \in \bar{s}\}$ and I_k, $k \in \bar{s}$, is defined as in Section 2.2. It is assumed that $\mathscr{M}$ has dimension greater than s. The fundamental result of this section is that if $\hat{\boldsymbol{\mu}}$ is the MLE of $\boldsymbol{\mu} \in \mathscr{M}$ for the Poisson model, then $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}^{(m)}$. This equation means that if one side exists, then the other side exists and the two sides are equal. This result is extremely useful since it implies that conditions for existence of MLE's under Poisson sampling also apply to multinomial sampling, and it is important in both numerical and algebraic work since it permits use of the relatively simple Poisson log likelihood in estimation problems involving multinomial sampling. This point is discussed by Haberman (1970 and 1972). Results of this section are related to those of Birch (1963), who provides a detailed analysis of the complete three-way table.

In order to examine maximum likelihood estimation for the multinomial case, the function

$$\hat{l}^{(m)}(\mathbf{n}, \boldsymbol{\mu}^*) = (\mathbf{n}, \boldsymbol{\mu}^*) - \sum_{k=1}^{s} (\mathbf{n}, \boldsymbol{\nu}^{(k)}) \log (\mathbf{m}(\boldsymbol{\mu}^*), \boldsymbol{\nu}^{(k)}) \tag{3.40}$$

defined for $\boldsymbol{\mu}^* \in \mathscr{M} - \mathscr{N}$ may be considered. By (2.24), to every $\boldsymbol{\mu}^* \in \mathscr{M} - \mathscr{N}$ corresponds a unique $\boldsymbol{\mu} \in \tilde{\mathscr{M}}$ such that $P_{\mathscr{M}-\mathscr{N}}\boldsymbol{\mu} = \boldsymbol{\mu}^*$. One may write $\boldsymbol{\mu}$ as $\mathbf{w}(\boldsymbol{\mu}^*)$. If $\hat{l}^{(m)}(\mathbf{n}, \boldsymbol{\mu}^*)$ has a maximum for $\boldsymbol{\mu}^* = \hat{\boldsymbol{\mu}}^*$, then $l^{(m)}(\mathbf{n}, \boldsymbol{\mu})$ has a maximum for $\boldsymbol{\mu} = \mathbf{w}(\hat{\boldsymbol{\mu}}^*)$. If $l^{(m)}(\mathbf{n}, \boldsymbol{\mu})$ has a maximum for $\boldsymbol{\mu} = \hat{\boldsymbol{\mu}}$, then $\hat{l}^{(m)}(\mathbf{n}, \boldsymbol{\mu}^*)$ has a maximum for $\boldsymbol{\mu}^* = P_{\mathscr{M}-\mathscr{N}}\hat{\boldsymbol{\mu}}$. Thus maximization of $l^{(m)}(\mathbf{n}, \boldsymbol{\mu})$ for $\boldsymbol{\mu} \in \mathscr{M}$ is equivalent to maximization of $\hat{l}^{(m)}(\mathbf{n}, \boldsymbol{\mu}^*)$ for $\boldsymbol{\mu}^* \in \mathscr{M} - \mathscr{N}$.

To examine the properties of the MLE $\hat{\boldsymbol{\mu}}^{(m)}$, it is necessary to examine the first and second differentials of $\hat{l}^{(m)}(\mathbf{n}, \boldsymbol{\mu}^*)$. Since

$$(3.41) \qquad \hat{l}^{(m)}(\mathbf{n}, \boldsymbol{\mu}^* + \boldsymbol{\nu}) = l(\mathbf{n}, \boldsymbol{\mu}^*) + (\mathbf{n}, \boldsymbol{\nu}) - \sum_{k=1}^{s} \frac{(\mathbf{n}, \boldsymbol{\nu}^{(k)})}{(m(\boldsymbol{\mu}^*), \boldsymbol{\nu}^{(k)})} \sum_{i \in I_k} m_i(\boldsymbol{\mu}^*)\nu_i^{(k)}\nu_i + o(\boldsymbol{\nu}),$$

it follows that if $\boldsymbol{\nu} \in \mathscr{M} - \mathscr{N}$, then

$$(3.42) \qquad d\hat{l}_{\boldsymbol{\mu}^*}^{(m)}(\mathbf{n}, \boldsymbol{\nu}) = (\boldsymbol{\nu}, \mathbf{n}) - \sum_{k=1}^{s} \frac{(\mathbf{n}, \boldsymbol{\nu}^{(k)})}{(\mathbf{m}(\boldsymbol{\mu}^*), \boldsymbol{\nu}^{(k)})} \sum_{i \in I_k} m_i(\boldsymbol{\mu}^*)\nu_i^{(k)}\nu_i .$$

If $\bar{\mathbf{m}}(\boldsymbol{\mu}^*) = \{e^{w_i(\mu^*)}\}$, then

$$(3.43) \qquad d\hat{l}_{\boldsymbol{\mu}^*}^{(m)}(\mathbf{n}, \boldsymbol{\nu}) = (\boldsymbol{\nu}, \mathbf{n} - \bar{\mathbf{m}}(\boldsymbol{\mu}^*)) .$$

In order to find the second differential $d^2\hat{l}_{\mu^*}^{(m)}(\mathbf{n}, \boldsymbol{\xi})(\boldsymbol{\nu})$, it is only necessary to note that

$$(3.44) \qquad d\hat{l}_{\boldsymbol{\mu}^*+\boldsymbol{\xi}}^{(m)}(\mathbf{n}, \boldsymbol{\nu}) = d\hat{l}_{\boldsymbol{\mu}^*}^{(m)}(\mathbf{n}, \boldsymbol{\nu}) - \sum_{k=1}^{s} \frac{(\mathbf{n}, \boldsymbol{\nu}^{(k)})}{(\mathbf{m}(\boldsymbol{\mu}^*), \boldsymbol{\nu}^{(k)})} \left\{ \sum_{i \in I_k} \nu_i \xi_i m_i(\boldsymbol{\mu}^*) - \frac{(\sum_{i \in I_k} m_i(\boldsymbol{\mu}^*)\nu_i)(\sum_{i \in I_k} m_i(\boldsymbol{\mu}^*)\xi_i)}{(\mathbf{m}(\boldsymbol{\mu}^*), \boldsymbol{\nu}^{(k)})} \right\} + o_{\boldsymbol{\xi}}(\nu) .$$

Thus

$$(3.45) \qquad d^2\hat{l}_{\mu^*}(\mathbf{n}, \boldsymbol{\xi})(\boldsymbol{\nu}) = -(\boldsymbol{\nu}, \bar{D}(\boldsymbol{\mu}^*)\boldsymbol{\xi}) + \sum_{k=1}^{s} \frac{(\boldsymbol{\nu}, \bar{D}(\boldsymbol{\mu}^*)\boldsymbol{\nu}^{(k)})(\boldsymbol{\xi}, \bar{D}(\boldsymbol{\mu}^*)\boldsymbol{\nu}^{(k)})}{(\boldsymbol{\nu}^{(k)}, \bar{D}(\boldsymbol{\mu}^*)\boldsymbol{\nu}^{(k)})}$$

where $\bar{D}(\boldsymbol{\mu}^*)\{x_i\} = \{\bar{m}_i(\boldsymbol{\mu}^*)x_i\}$. If $\bar{P}_{\mathscr{N}}(\mu^*)$ is the orthogonal projection on $\mathscr{N}$ relative to the inner product $((\cdot, \cdot))$ defined by $((\mathbf{x}, \mathbf{y})) = (\mathbf{x}, \bar{D}(\boldsymbol{\mu}^*)\mathbf{y})$ for $\mathbf{x} \in R^I$ and $\mathbf{y} \in R^I$, then

$$(3.46) \qquad \begin{aligned} d^2\hat{l}_{\mu}^{(m)}(\mathbf{n}, \boldsymbol{\xi})(\boldsymbol{\nu}) &= -(\boldsymbol{\nu}, \bar{D}(\boldsymbol{\mu}^*)[I - \bar{P}_{\mathscr{N}}(\boldsymbol{\mu}^*)]\boldsymbol{\xi}) \\ &= -(\bar{Q}_{\mathscr{N}}(\boldsymbol{\mu}^*)\boldsymbol{\nu}, \bar{D}(\boldsymbol{\mu}^*)\bar{Q}_{\mathscr{N}}(\boldsymbol{\mu}^*)\boldsymbol{\xi}) . \end{aligned}$$

Here $\bar{Q}_{\mathscr{N}}(\boldsymbol{\mu}^*) = I - \bar{P}_{\mathscr{N}}(\boldsymbol{\mu}^*)$. Since $-(\bar{Q}_{\mathscr{N}}(\boldsymbol{\mu}^*)\boldsymbol{\nu}, \bar{D}(\boldsymbol{\mu}^*)\bar{Q}_{\mathscr{N}}(\boldsymbol{\mu}^*)\boldsymbol{\nu}) \leqq 0$ for $\boldsymbol{\nu} \in \mathscr{M} - \mathscr{N}$, $\hat{l}^{(m)}(\mathbf{n}, \boldsymbol{\mu}^*)$is concave. If $-(\bar{Q}_{\mathscr{N}}(\boldsymbol{\mu}^*)\boldsymbol{\nu}, \bar{D}(\boldsymbol{\mu}^*)\bar{Q}_{\mathscr{N}}(\boldsymbol{\mu}^*)\boldsymbol{\nu}) = 0$, then $\bar{Q}_{\mathscr{N}}(\boldsymbol{\mu}^*)\boldsymbol{\nu} = \mathbf{0}$. Therefore, $\boldsymbol{\nu} \in \mathscr{N}$. Since $\boldsymbol{\nu} \in \mathscr{M} - \mathscr{N}$, $\boldsymbol{\nu} = 0$. Thus $\hat{l}^{(m)}(\mathbf{n}, \boldsymbol{\mu}^*)$ is strictly concave for $\boldsymbol{\mu}^* \in \mathscr{M} - \mathscr{N}$.

The fundamental result for multinomial models now follows:

THEOREM 3.4. *If $\hat{\boldsymbol{\mu}}^{(m)}$ is the* MLE *for a multinomial model for which $\boldsymbol{\mu} \in \mathscr{M}$ and if $\hat{\boldsymbol{\mu}}$ is the corresponding estimate for a Poisson model for which $\boldsymbol{\mu} \in \mathscr{M}$, then $\hat{\boldsymbol{\mu}}^{(m)} = \hat{\boldsymbol{\mu}}$, in the sense that when one side of the equation exists, then the other side exists and the two sides are equal.*

PROOF. Suppose $\hat{\boldsymbol{\mu}}$ exists. Then for $\boldsymbol{\nu} \in \mathscr{M}$,

$$(3.47) \qquad (\boldsymbol{\nu}, \mathbf{n} - \hat{\mathbf{m}}) = 0 .$$

Since $(\mathbf{n}, \boldsymbol{\nu}^{(k)}) = (\hat{\mathbf{m}}, \boldsymbol{\nu}^{(k)})$ for $k \in \bar{s}$, $\hat{\boldsymbol{\mu}} \in \mathscr{M}$. If $\hat{\boldsymbol{\mu}}^* = P_{\mathscr{M}-\mathscr{N}}\hat{\boldsymbol{\mu}}$, then for any $\boldsymbol{\nu} \in \mathscr{M} - \mathscr{N}$,

$$(3.48) \qquad d\hat{l}_{\hat{\boldsymbol{\mu}}^*}^{(m)}(\mathbf{n}, \boldsymbol{\nu}) = (\boldsymbol{\nu}, \mathbf{n} - \hat{\mathbf{m}}) = 0 .$$

Therefore, $\hat{l}^{(m)}(\mathbf{n}, \boldsymbol{\mu}^*)$ has a critical point at $\hat{\boldsymbol{\mu}}^*$. Since $\hat{l}^{(m)}(\mathbf{n}, \boldsymbol{\mu}^*)$ is strictly concave, this critical point, a maximum, is the only point $\boldsymbol{\mu}^*$ for which $d\hat{l}^{(m)}_{\boldsymbol{\mu}^*}(\mathbf{n}, \boldsymbol{\nu})$ is 0 for all $\boldsymbol{\nu} \in \mathscr{M} - \mathscr{N}$. Thus $\hat{\boldsymbol{\mu}}^{(m)} = \mathbf{w}(\hat{\boldsymbol{\mu}}^*) = \hat{\boldsymbol{\mu}}$.

On the other hand, suppose that $\hat{\boldsymbol{\mu}}^{(m)}$ exists. Then for $\boldsymbol{\nu} \in \mathscr{M} - \mathscr{N}$,

$$(\boldsymbol{\nu}, \mathbf{n} - \hat{\mathbf{m}}^{(m)}) = 0\,, \tag{3.49}$$

where $\hat{\mathbf{m}}^{(m)} = \{e^{\hat{\mu}_i^{(m)}}\}$. If $k \in \bar{s}$, then $(\boldsymbol{\nu}^{(k)}, \hat{\mathbf{m}}^{(m)}) = (\boldsymbol{\nu}^{(k)}, \mathbf{n})$. Thus for any $\boldsymbol{\nu} \in \mathscr{M}$,

$$(\boldsymbol{\nu}, \mathbf{n} - \hat{\mathbf{m}}^{(m)}) = 0\,. \tag{3.50}$$

By Theorem 3.1, $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}^{(m)}$. □

Since Theorem 3.4 holds, the results concerning necessary and sufficient conditions for existence of MLE's under Poisson sampling also apply to multinomial sampling.

EXAMPLE 3.4. In the independence model for the $r \times c$ contingency table of Example 2.1, it is well known that whenever $n_{i+} > 0$ for $i \in \bar{r}$ and $n_{+j} > 0$ for $j \in \bar{c}$, then

$$\hat{m}_{ij}^{(m)} = n_{i+} n_{+j}/N\,. \tag{3.51}$$

This result may be verified by use of Theorems 3.1 and 3.4, for whenever $n_{i+} > 0$ for $i \in \bar{r}$ and $n_{+j} > 0$ for $j \in \bar{c}$, $\{\log(n_{i+} n_{+j}/N)\} \in \mathscr{M}$ and

$$\begin{aligned} P_{\mathscr{M}} n_{i+} n_{+j}/N &= \left\{\frac{1}{r} n_{i+} + \frac{1}{c} n_{+j} - \frac{1}{rc} N\right\} \\ &= P_{\mathscr{M}} \mathbf{n} \end{aligned} \tag{3.52}$$

(see Kruskal (1968)). On the other hand, if for some $i' \in \bar{r}$, $n_{i'+} = 0$, then

$$(\mathbf{n}, \mathbf{x}) = 0 \tag{3.53}$$

for

$$\begin{aligned} x_{ij} &= -1 \qquad \text{if} \quad i = i'\,, \\ &= 0 \qquad \text{otherwise}\,. \end{aligned} \tag{3.54}$$

Since $\mathbf{x} \in \mathscr{M}$, $\mathbf{x} \neq \mathbf{0}$, and $x_{ij} \leqq 0$ for $(i, j) \in \bar{r} \times \bar{c}$, Theorems 3.3 and 3.4, together with (3.35), imply that $\hat{\mathbf{m}}^{(m)}$ does not exist. A similar argument shows that $\hat{\mathbf{m}}^{(m)}$ does not exist if $n_{+j'} = 0$ for some $j' \in \bar{c}$. Thus $\hat{\mathbf{m}}^{(m)}$ exists if and only if $n_{i+} > 0$ for $i \in \bar{r}$ and $n_{+j} > 0$ for $i \in \bar{c}$. If $\hat{\mathbf{m}}^{(m)}$ exists, then it is given by (3.51).

EXAMPLE 3.5. Consider the quantal response model of Examples 2.2 and 2.6. The linear manifold $\mathscr{M}$ is the span of $\mathbf{x}$, $\mathbf{y}$, and $\boldsymbol{\nu}^{(k)}$, $k \in \bar{r}$, where these vectors are defined as in Example 2.6. Theorems 3.1 and 3.4 imply that if $\hat{\mathbf{m}}^{(m)}$ exists, then

$$\begin{aligned} (\mathbf{x}, \hat{\mathbf{m}}^{(m)}) &= \hat{m}_{+1}^{(m)} - \hat{m}_{+2}^{(m)} \\ &= n_{+1} - n_{+2}\,, \end{aligned} \tag{3.55}$$

$$\begin{aligned} (\mathbf{y}, \hat{\mathbf{m}}^{(m)}) &= \textstyle\sum_{j=1}^{r} t_j(\hat{m}_{j1}^{(m)} - \hat{m}_{j2}^{(m)}) \\ &= \textstyle\sum_{j=1}^{r} t_j(n_{j1} - n_{j2})\,, \end{aligned} \tag{3.56}$$

and

$$(\boldsymbol{\nu}^{(j)}, \hat{\mathbf{m}}^{(m)}) = \hat{m}_{j+}^{(m)} = n_{j+} = N_j \tag{3.57}$$

for $j \in \bar{r}$. Since (3.57) implies that $n_{j2} = N_j - n_{j1}$ and $\hat{m}_{j2}^{(m)} = N_j - \hat{m}_{j1}^{(m)}$, (3.55) and (3.56) may be replaced by the equations

$$\hat{m}_{+1}^{(m)} = n_{+1} \tag{3.58}$$

and

$$\sum_{j=1}^{r} t_j \hat{m}_{j1}^{(m)} = \sum_{j=1}^{r} t_j n_{j1} \,, \tag{3.59}$$

where $\hat{m}_{j1}^{(m)}$ may be written

$$N_j/(1 + e^{-(\hat{\alpha}^{(m)} + \hat{\beta}^{(m)} t_j)}) \tag{3.60}$$

and $\hat{\alpha}^{(m)}$ and $\hat{\beta}^{(m)}$ are the MLE's of α and β, respectively. These likelihood equations are consistent with those given in standard works on quantal response such as Finney (1952). It should be noted that (3.55), (3.56), and (3.57) are still the likelihood equations if the n_{jk}, $1 \leqq j \leqq r$, $1 \leqq k \leqq 2$, are independent Poisson random variables or if $\mathbf{n}$ has been obtained from a single multinomial sample of size $N > 0$.

To find necessary and sufficient conditions for the existence of $\hat{\mathbf{m}}^{(m)}$, order the t_j so that $t_j < t_{j'}$ if $j < j'$. Given this condition, $\hat{\mathbf{m}}^{(m)}$ exists if and only if for no $j' \in \bar{r}$ is it the case that either (a) $n_{j1} = 0$ for $j < j'$ and $n_{j2} = 0$ for $j > j'$ or (b) $n_{j1} = 0$ for $j > j'$ and $n_{j2} = 0$ for $j < j'$.

If $\hat{\mathbf{m}}^{(m)}$ does not exist, Theorems 3.3 and 3.4 imply that for some $\boldsymbol{\mu} \in \mathscr{M}$, $\boldsymbol{\mu} \neq 0$, $\mu_{jk} \leqq 0$ for $j \in \bar{r}$ and $k \in \bar{2}$, and $(\mathbf{n}, \boldsymbol{\mu}) = 0$. If $\boldsymbol{\mu} \in \mathscr{M}$, then for some a, b, and $\{c_j : j \in \bar{r}\}$,

$$\boldsymbol{\mu} = a\mathbf{x} + b\mathbf{y} + \sum_{j=1}^{r} c_j \boldsymbol{\nu}^{(j)} \,. \tag{3.61}$$

Since $N_j > 0$ for each $j \in \bar{r}$, either $n_{j1} > 0$ or $n_{j2} > 0$. If $n_{j1} > 0$, $\mu_{j1} = 0$, $c_j = -(a + bt_j)$, and $\mu_{j2} = -2(a + bt_j) \leqq 0$. If $n_{j2} > 0$, $\mu_{j2} = 0$, $c_j = a + bt_j$, and $\mu_{j1} = 2(a + bt_j) \leqq 0$.

Suppose $A = \{j \in \bar{r} : n_{j1} > 0 \text{ and } n_{j2} = 0\}$, $B = \{j \in \bar{r} : n_{j2} > 0 \text{ and } n_{j1} = 0\}$, and $C = \{j \in \bar{r} : n_{j1} > 0 \text{ and } n_{j2} > 0\}$. Then A, B and C are disjoint sets with union $A \cup B \cup C = \bar{r}$. Thus $a + bt_j \geqq 0$ for $j \in A$, $a + bt_j \leqq 0$ for $j \in B$, and $a + bt_j = 0$ for $j \in C$. If C has 2 or more elements, then since $t_j \neq t_{j'}$ if $j \neq j'$, $a = b = 0$ and $\boldsymbol{\mu} = \mathbf{0}$, a contradiction. Thus C has no more than 1 element. There are now 3 possibilities: $b > 0$, $b = 0$, or $b < 0$. If $b > 0$, then for some $j' \in \bar{r}$, $a + bt_j > 0$ for $j > j'$ and $a + bt_j < 0$ for $j < j'$. Thus $n_{j1} = 0$ for $j < j'$ and $n_{j2} = 0$ for $j > j'$. Similarly, if $b < 0$, then for some $j' \in \bar{r}$, $n_{j1} = 0$ for $j > j'$ and $n_{j2} = 0$ for $j < j'$. If $b = 0$, then in order that $\boldsymbol{\mu} \neq \mathbf{0}$, $a > 0$ or $a < 0$. If $a > 0$, $n_{j2} = 0$ for all $j \in \bar{r}$, while if $a < 0$, $n_{j1} = 0$ for all $j \in \bar{r}$. Thus (a) or (b) holds for some $j' \in \bar{r}$.

On the other hand, suppose that for some $j' \in \bar{r}$, (a) or (b) holds. Without loss of generality, suppose that (a) holds. Let $a = -t_{j'}$, $b = 1$, and $c_j = -|a + bt_j|$. Suppose $\boldsymbol{\mu}$ satisfies (3.61). Then $\boldsymbol{\mu} \in \mathscr{M}$, $\mu_{jk} \leqq 0$ for $j \in \bar{r}$ and $k \in \bar{2}$, $\boldsymbol{\mu} \neq \mathbf{0}$, and $(\mathbf{n}, \boldsymbol{\mu}) = 0$. Therefore, $\hat{\mathbf{m}}^{(m)}$ does not exist. These conditions have been used in the case $r = 3$ by Silverstone (1957).

4. Conclusion. In this paper, a general log-linear model for use with frequency data has been proposed. This model has been applied to logit analysis and the analysis of factorial tables. Discussion in this paper has emphasized construction of complete minimal sufficient statistics and likelihood equations. Future papers will consider computation of maximum likelihood estimates and determination of asymptotic properties for these estimates.

REFERENCES

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

[2] BISHOP, Y. M. M. (1969). Full contingency tables, logits, and split contingency tables. *Biometrics* **25** 383–400.

[3] BISHOP, Y. M. M. and FIENBERG, S. E. (1969). Incomplete two-dimensional contingency tables. *Biometrics* **25** 119–128.

[4] BLACKWELL, D. and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.

[5] DARROCH, J. N. (1962). Interaction in multifactor contingency tables. *J. Roy. Statist. Soc. Ser. B* **24** 251–263.

[6] DYKE, G. V. and PATTERSON, H. D. (1952). Analysis of factorial arrangements when the data are proportions. *Biometrics* **8** 1–12.

[7] FIENBERG, S. E. (1970). Quasi-independence and maximum likelihood estimation in incomplete contingency tables. *J. Amer. Statist. Assoc.* **65** 1610–1616.

[8] FIENBERG, S. E. (1972). The analysis of incomplete multiway contingency tables. *Biometrics* **28** 177–202.

[9] FINNEY, D. J. (1952). *Probit Analysis.* Cambridge Univ. Press.

[10] GOODMAN, L. A. (1968). The analysis of cross-classified data: independence, quasi-independence and interactions in contingency tables with or without missing entries. *J. Amer. Statist. Assoc.* **63** 1091–1131.

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

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

[13] HABERMAN, S. J. (1972). *The Analysis of Frequency Data.* Submitted to IMS Monograph Series.

[14] HALMOS, P. R. (1958). *Finite Dimensional Vector Spaces.* Van Nostrand, Princeton.

[15] KRUSKAL, W. H. (1968). Notes on analysis of variance. Unpublished MS, Univ. of Chicago.

[16] SILVERSTONE, H. (1957). Estimating the logistic curve. *J. Amer. Statist. Assoc.* **52** 567–577.

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