The Annals of Statistics
1977, Vol. 5, No. 6, 1148-1169

LOG-LINEAR MODELS AND FREQUENCY TABLES WITH SMALL EXPECTED CELL COUNTS[1]

BY SHELBY J. HABERMAN

University of Chicago

In the case of frequency data, traditional discussions such as Rao (1973, pages 355-363, 391-412) consider asymptotic properties of maximum likelihood estimates and chi-square statistics under the assumption that all expected cell frequencies become large. If log-linear models are applied, these asymptotic properties may remain applicable if the sample size is large and the number of cells in the table is large, even if individual expected cell frequencies are small. Conditions are provided for asymptotic normality of linear functionals of maximum-likelihood estimates of log-mean vectors and for asymptotic chi-square distributions of Pearson and likelihood ratio chi-square statistics.

1. Introduction. Frequency tables are commonly encountered in which the total number of observations is large and the number of cells is finite, but so many cells are present that individual cell frequencies are small. If log-linear models are used with such tables, it is necessary to have some knowledge of the usefulness of customary asymptotic results for maximum likelihood estimates and chi-square tests.

Common treatments of asymptotic theory do not apply to this type of frequency table. For example, Rao's (1973, pages 355-363, 391-412) results apply to a multinomial vector $\mathbf{n}$ with elements n_i, $1 \leqq i \leqq k$. The sample size N is assumed to approach infinity, while the number of elements k of $\mathbf{n}$ and the cell probabilities p_i, $1 \leqq i \leqq k$, all remain fixed. Thus the expected cell counts Np_i, $1 \leqq i \leqq k$, all approach infinity as N approaches infinity. Except in the case of multinomial response models, Haberman's (1974, pages 74-122) discussion of asymptotic properties of log-linear models applies to a frequency table $\mathbf{n} = \{n_i : i \in I\}$ with elements n_i indexed by elements of a finite nonempty set I. Each count n_i has expected value m_i, and for some positive c_i, $i \in I$, it is assumed that

$$N^{-1}m_i \to c_i\,, \qquad i \in I\,,$$

and

$$N = \sum_{i \in I} m_i \to \infty\,.$$

Thus each m_i goes to infinity as N goes to infinity.

To develop appropriate asymptotic theory for frequency tables with small

Received February 1975; revised January 1977.

[1] Support for this research has been provided in part by National Science Foundation Research Grants No. SOC 72-05228 A04 and No. MCS 72-04364 A04.

AMS 1970 *subject classifications*. Primary 62E20; Secondary 62F05, 62F10, 62F25.

Key words and phrases. Contingency tables, log-linear models, maximum likelihood, chi-square tests, asymptotic properties.

expected cell counts, it is necessary to consider a sequence of frequency tables with possibly varying numbers of elements. For this purpose, for each nonnegative integer t, consider a frequency table $\mathbf{n}_t$ with cell counts n_{it}, $i \in I_t$, where I_t is a finite nonempty index set. Let n_{it} have positive expected value m_{it}, and let $\mu_{it} = \log m_{it}$. For each t, a log-linear model is considered in which the vector $\boldsymbol{\mu}_t$ with elements μ_{it}, $i \in I_t$, is assumed to belong to a linear manifold $\mathscr{M}_t$. Nontrivial asymptotic results developed in this paper will require that the sum

$$N_t = \sum_{i \in I_t} m_{it}$$

approach infinity as t approaches infinity. It will not be necessary that

$$\min_{i \in I_t} m_{it}$$

approach infinity as t approaches infinity. However, restrictions will be imposed on the sequence $\mathbf{m}_t = \{m_{it} : i \in I_t\}$, $t \geqq 0$, of expected values and on the sequence $\mathscr{M}_t$, $t \geqq 0$, of linear manifolds. These restrictions depend on the sampling procedures used to generate the observed tables $\mathbf{n}_t$.

As in Haberman (1973, 1974), the table $\mathbf{n}_t$ may satisfy a Poisson or a multinomial sampling model. In the Poisson sampling model, the counts n_{it}, $i \in I_t$, are mutually independent and have Poisson distributions. In the multinomial sampling model, $\mathbf{n}_t$ consists of $r_t \geqq 1$ independently distributed subtables $\mathbf{n}_{kt}$, $1 \leqq k \leqq r_t$. For $1 \leqq k \leqq r_t$, the subtable $\mathbf{n}_{kt}$ includes the counts n_{it}, $i \in J_{kt}$, and $\mathbf{n}_{kt}$ has a multinomial distribution with sample size N_{kt}. In this sampling model, the J_{kt}, $1 \leqq k \leqq r_t$, are disjoint sets with union I_t.

The sampling model may be characterized by a linear manifold $\mathscr{N}_t$ consisting of all vectors $\mathbf{x} = \{x_i : i \in I_t\}$ such that

$$\langle \mathbf{x}, \mathbf{n}_t \rangle_t = \sum_{i \in I_t} x_i n_{it} = \sum_{i \in I_t} x_i m_{it} = \langle \mathbf{x}, \mathbf{m}_t \rangle_t$$

for any observed table $\mathbf{n}_t$ consistent with the sampling model. In the case of Poisson sampling, each count n_{it} may take on any nonnegative integral value, so that $\mathscr{N}_t$ consists of the single vector $\mathbf{0}$ with all elements zero. In the multinomial sampling case, the table $\mathbf{n}_t$ is constrained by the linear equations

$$\sum_{i \in J_{kt}} n_{it} = \sum_{i \in J_{kt}} m_t = N_k\,, \qquad 1 \leqq k \leqq r_t\,.$$

If

$$\begin{aligned} \nu_{ikt} &= 1\,, \quad i \in J_{kt}\,, \\ &= 0\,, \quad i \in I_t - J_{kt}\,, \end{aligned}$$

then

$$\langle \boldsymbol{\nu}_{kt}, \mathbf{n} \rangle_t = \langle \boldsymbol{\nu}_{kt}, \mathbf{m}_t \rangle_t\,, \qquad 1 \leqq k \leqq r_t\,.$$

Thus $\mathscr{N}_t$ is the span of the vectors $\boldsymbol{\nu}_{kt}$, $1 \leqq k \leqq r_t$. In log-linear models considered in this paper, $\mathscr{N}_t$ is assumed to be included in $\mathscr{M}_t$.

Let $\hat{\boldsymbol{\mu}}_t$ be the maximum likelihood estimate of $\boldsymbol{\mu}_t$ under the model $\boldsymbol{\mu}_t \in \mathscr{M}_t$. Let $\hat{\boldsymbol{\mu}}_t'$ be the maximum likelihood estimate of $\boldsymbol{\mu}_t$ under the alternate model $\boldsymbol{\mu}_t \in \mathscr{M}_t'$, where $\mathscr{M}_t'$ is a linear manifold such that $\mathscr{M}_t \subset \mathscr{M}_t'$. Let X_t^2 be the Pearson chi-square statistic and let L_t^2 be the likelihood ratio chi-square statistic for the null hypothesis that $\boldsymbol{\mu}_t \in \mathscr{M}_t$ against the alternative hypothesis that $\boldsymbol{\mu}_t \in \mathscr{M}_t'$. As noted in Haberman (1973), $\hat{\boldsymbol{\mu}}_t$ and $\hat{\boldsymbol{\mu}}_t'$ are uniquely determined if they exist by the equations

$$(1.1) \qquad \langle \hat{\mathbf{m}}_t, \mathbf{x}\rangle_t = \sum_{i \in I_t} \hat{m}_{it} x_i = \sum_{i \in I_t} n_{it} x_i = \langle \mathbf{n}_t, \mathbf{x}\rangle_t \,, \qquad \mathbf{x} \in \mathscr{M}_t \,,$$

and

$$(1.2) \qquad \langle \hat{\mathbf{m}}_t', \mathbf{x}\rangle_t = \langle \mathbf{n}_t, \mathbf{x}\rangle_t \,, \qquad \mathbf{x} \in \mathscr{M}_t' \,,$$

where $\hat{m}_{it} = \exp \hat{\mu}_{it}$ and $\hat{m}_{it}' = \exp \hat{\mu}_{it}'$ for $i \in I_t$. Given $\hat{\boldsymbol{\mu}}_t$, $\hat{\boldsymbol{\mu}}_t'$, $\hat{\mathbf{m}}_t$, and $\hat{\mathbf{m}}_t'$, one has

$$(1.3) \qquad X_t^2 = \sum_{i \in I_t} (\hat{m}_{it}' - \hat{m}_{it})^2 / \hat{m}_{it}$$

and

$$(1.4) \qquad L_t^2 = 2 \sum_{i \in I_t} [n_{it} \log (\hat{m}_{it}'/\hat{m}_{it}) + \hat{m}_{it} - \hat{m}_{it}'] \,.$$

If the unit vector $\mathbf{e}_t$ is in $\mathscr{M}_t$, where all elements of $\mathbf{e}_t$ are 1, then

$$L_t^2 = 2 \sum_{i \in I_t} n_{it} \log (\hat{m}_{it}'/\hat{m}_{it}) = 2\langle \mathbf{n}_t, \hat{\boldsymbol{\mu}}_t' - \hat{\boldsymbol{\mu}}_t \rangle_t \,.$$

This paper examines asymptotic properties of $\hat{\boldsymbol{\mu}}_t$, $\hat{\boldsymbol{\mu}}_t'$, X_t^2, and L_t^2 as $t \to \infty$.

In Section 2, conditions are presented under which linear functionals of the maximum likelihood estimates $\hat{\boldsymbol{\mu}}_t$ are asymptotically normal. In Section 3, conditions are presented under which the test statistics X_t^2 and L_t^2 have asymptotic chi-square distributions and are asymptotically equivalent. Sections 4 and 5 provide proofs of results in Section 2, while Section 6 provides proofs of results in Section 3.

2. Asymptotic properties of $\hat{\boldsymbol{\mu}}_t$. Description of asymptotic properties of $\hat{\boldsymbol{\mu}}_t$ is complicated by the fact that $\mathscr{N}_t$, $\mathscr{M}_t$, and I_t all depend on t. Consequently, it is not in general possible to say that a suitably normalized version of $\hat{\boldsymbol{\mu}}_t$ converges in distribution to some random vector $\mathbf{Y}$. Nevertheless, asymptotic properties of sequences $\{\gamma_t(\hat{\boldsymbol{\mu}}_t)\}$ can be considered, where for $t \geqq 0$, γ_t is a linear functional on $\mathscr{M}_t$. In the case of multinomial sampling, linear functionals of interest depend on probabilities $p_{it} = m_{it}/N_{kt}$, $i \in J_{kt}$, $1 \leqq k \leqq r_t$, rather than on sample sizes N_{kt}, $1 \leqq k \leqq r_t$. Therefore, the assumption will be made that

$$(2.1) \qquad \gamma_t(\mathbf{x}) = 0 \,, \qquad \mathbf{x} \in \mathscr{N}_t \,.$$

To avoid trivial cases in which $\gamma_t(\hat{\boldsymbol{\mu}}_t)$ is identically zero, it is assumed that $\gamma_t(\mathbf{x}) \neq 0$ for some $\mathbf{x} \in \mathscr{M}_t$. Given (2.1), it must also be the case that $\mathscr{N}_t \neq \mathscr{M}_t$. In this section, conditions are given under which $\hat{\gamma}_t = \gamma_t(\hat{\boldsymbol{\mu}}_t)$ has an asymptotically normal distribution with asymptotic mean $\gamma_{t0} = \gamma_t(\boldsymbol{\mu}_t)$ and asymptotic variance

$$(2.2) \qquad \sigma^2(\hat{\gamma}_t) = \sup\{|\gamma_t(\mathbf{x})|^2 : \mathbf{x} \in \mathscr{M}_t, \sum_{i \in I_t} x_i^2 m_{it} = 1\} \,;$$

so that as $t \to \infty$, the standardized ratio

$$(\hat{\gamma}_t - \gamma_{t0})/\sigma(\hat{\gamma}_t)$$

converges in distribution to the standard normal distribution $N(0, 1)$. These conditions also ensure that the estimated asymptotic variance

$$\hat{\sigma}^2(\hat{\gamma}_t) = \sup\{|\gamma_t(\mathbf{x})| : \mathbf{x} \in \mathscr{M}_t, \sum_{i \in I_t} x_i^2 \hat{m}_{it} = 1\} \tag{2.3}$$

is a consistent estimate of the asymptotic variance $\sigma^2(\hat{\gamma}_t)$ in the sense that the ratio $\hat{\sigma}^2(\hat{\gamma}_t)/\sigma^2(\hat{\gamma}_t)$ converges in probability to 1 as $t \to \infty$. If $0 < \alpha < 1$ and $Z_{\alpha/2}$ is the upper-$\alpha/2$ point of the standard normal distribution, then as $t \to \infty$, the probability approaches $1 - \alpha$ that

$$\hat{\gamma}_t - Z_{\alpha/2}\hat{\sigma}(\hat{\gamma}_t) \leqq \gamma_{t0} \leqq \hat{\gamma}_t + Z_{\alpha/2}\hat{\sigma}(\hat{\gamma}_t) .$$

Thus as approximate level-$(1 - \alpha)$ confidence interval for γ_{t0} has lower bound

$$\hat{\gamma}_t - Z_{\alpha/2}\hat{\sigma}(\hat{\gamma}_t)$$

and upper bound

$$\hat{\gamma}_t + Z_{\alpha/2}\hat{\sigma}(\hat{\gamma}_t) .$$

The formulas for the asymptotic variance $\sigma^2(\hat{\gamma}_t)$ and the estimated asymptotic variance $\hat{\sigma}^2(\hat{\gamma}_t)$ are consistent with those obtained by Haberman (1974, pages 75–81), despite differences in appearance. If for some $\mathbf{c}_t$,

$$\gamma_t(\mathbf{x}) = \langle \mathbf{c}_t, \mathbf{x} \rangle_t , \qquad \mathbf{x} \in \mathscr{M}_t ,$$

then $\sigma^2(\hat{\gamma}_t)$ may be computed in the following manner. Let

$$[\mathbf{x}, \mathbf{y}]_t = \sum_{i \in I_t} x_i y_i m_{it} , \qquad \mathbf{x}, \mathbf{y} \in R^{I_t} , \tag{2.4}$$

where R^{I_t} consists of all vectors $\mathbf{x}$ with real coordinates x_i, $i \in I_t$. Let P_t be the projection on $\mathscr{M}_t$ orthogonal with respect to $[\cdot, \cdot]_t$. Let D_t be the linear transformation on R^{I_t} such that

$$D_t \mathbf{x} = \{x_i m_{it} : i \in I_t\} , \qquad \mathbf{x} \in R^{I_t} . \tag{2.5}$$

Then

$$\sigma^2(\hat{\gamma}_t) = \langle \mathbf{c}_t, P_t D_t^{-1} \mathbf{c}_t \rangle_t . \tag{2.6}$$

If γ_t satisfies (2.1), then

$$\langle \mathbf{c}_t, \mathbf{x} \rangle_t = 0 , \qquad \mathbf{x} \in \mathscr{N}_t . \tag{2.7}$$

Given (2.7), the right-hand side of (2.6) is the expression for the asymptotic variance of $\hat{\gamma}_t = \langle \mathbf{c}_t, \hat{\boldsymbol{\mu}}_t \rangle_t$ found in Haberman (1974, pages 80–81).

To verify (2.6), let $||\cdot||_t$ be the norm on R^{I_t} such that

$$||\mathbf{x}||_t^2 = [\mathbf{x}, \mathbf{x}]_t = \sum_{i \in I_t} x_i^2 m_{it} , \qquad \mathbf{x} \in R^{I_t} . \tag{2.8}$$

Note that P_t and D_t satisfy the equations

$$[P_t \mathbf{x}, \mathbf{y}]_t = [\mathbf{x}, \mathbf{y}]_t , \qquad \mathbf{x} \in R^{I_t}, \mathbf{y} \in \mathscr{M}_t , \tag{2.9}$$

and

$$\langle \mathbf{x}, \mathbf{y} \rangle_t = [D_t^{-1} \mathbf{x}, \mathbf{y}]_t , \qquad \mathbf{x}, \mathbf{y} \in R^{I_t} .$$

Thus

$$\gamma_t(\mathbf{x}) = [D_t^{-1}\mathbf{c}_t, \mathbf{x}]_t$$
$$= [P_t D_t^{-1}\mathbf{c}_t, \mathbf{x}]_t, \qquad \mathbf{x} \in \mathscr{M}_t.$$

Since $\gamma_t(\mathbf{x}) \neq 0$ for some $\mathbf{x} \in \mathscr{M}_t$, $||P_t D_t^{-1}\mathbf{c}_t||_t > 0$. By Schwartz's inequality,

$$|\gamma_t(\mathbf{x})| \leqq ||P_t D_t^{-1}\mathbf{c}_t||_t$$

whenever $\mathbf{x} \in \mathscr{M}_t$ and

$$||\mathbf{x}||_t = \sum_{i \in I_t} x_i^2 m_{it} = 1.$$

Equality is achieved if

$$\mathbf{x} = P_t D_t^{-1}\mathbf{c}_t / ||P_t D_t^{-1}\mathbf{c}_t||_t.$$

Thus

$$\sigma^2(\hat{\gamma}_t) = ||P_t D_t^{-1}\mathbf{c}_t||_t^2$$
$$= [P_t D_t^{-1}\mathbf{c}_t, P_t D_t^{-1}\mathbf{c}_t]_t$$
$$= [D_t^{-1}\mathbf{c}_t, P_t D_t^{-1}\mathbf{c}_t]_t$$
$$= \langle \mathbf{c}_t, P_t D_t^{-1}\mathbf{c}_t \rangle_t.$$

Similar arguments may be used to relate (2.3) to formulas of Haberman (1974, page 81) for the estimated asymptotic variance.

Conditions required for the asymptotic results described in this section depend on the orthogonal complement $\mathscr{S}_t$ of $\mathscr{N}_t$ relative to the linear manifold $\mathscr{M}_t$ and the inner product $[\cdot, \cdot]_t$. Thus

$$\mathscr{S}_t = \{\mathbf{y} \in \mathscr{M}_t : [\mathbf{x}, \mathbf{y}]_t = 0, \mathbf{x} \in \mathscr{N}_t\}. \tag{2.10}$$

The dimension a_t of $\mathscr{S}_t$ is $\dim \mathscr{M}_t - \dim \mathscr{N}_t$. Under the Poisson sampling model, $\mathscr{S}_t = \mathscr{M}_t$ and $a_t = \dim \mathscr{M}_t$. Under the multinomial sampling model, $a_t = \dim \mathscr{M}_t - r_t$.

Much of the analysis of this paper depends on the observation that there is a one-to-one correspondence between elements of $\mathscr{S}_t$ and the set $\mathscr{M}_{t*}$ of $\mathbf{x}$ in $\mathscr{M}_t$ such that

$$\sum_{i \in I_t} y_i \exp x_i = \sum_{i \in I_t} y_i m_{it}, \qquad \mathbf{y} \in \mathscr{N}_t.$$

This correspondence is important since $\hat{\boldsymbol{\mu}}_t$ is in $\mathscr{M}_{t*}$. To verify this correspondence, let $\boldsymbol{\mu}_t^*(\mathbf{x})$ and $\mathbf{m}_t^*(\mathbf{x})$ be defined for $\mathbf{x} \in \mathscr{M}_t$ so that

$$\boldsymbol{\mu}_t^*(\mathbf{x}) - \mathbf{x} \in \mathscr{N}_t,$$
$$\langle \mathbf{y}, \mathbf{m}_t^*(\mathbf{x}) \rangle_t = \langle \mathbf{y}, \mathbf{m}_t \rangle_t, \qquad \mathbf{y} \in \mathscr{N}_t,$$

and

$$m_{it}^*(\mathbf{x}) = \exp \mu_{it}^*(\mathbf{x}), \qquad i \in I_t.$$

Under the Poisson sampling model, $\boldsymbol{\mu}_t^*(\mathbf{x}) = \mathbf{x}$ and $m_{it}^*(\mathbf{x}) = \exp x_i$, $i \in I_t$. Under the multinomial sampling model,

$$m_{it}^*(\mathbf{x}) = N_{kt} \exp x_i / \sum_{j \in J_{kt}} \exp x_j, \qquad i \in J_{kt},\ 1 \leqq k \leqq r_t.$$

On the other hand, if $\mathbf{x} \in \mathscr{M}_t$ and

$$\sum_{i \in I_t} y_i \exp x_i = \sum_{i \in I_t} y_i m_{it}, \qquad \mathbf{y} \in \mathscr{N}_t,$$

then there is a unique element $\mathbf{z} \in \mathscr{S}_t$ such that $\mathbf{x} = \boldsymbol{\mu}^*(\mathbf{z})$ and

$$\exp x_i = m_{it}^*(\mathbf{z}) , \qquad i \in I_t .$$

If Q_t is the projection on $\mathscr{N}_t$ orthogonal with respect to $[\cdot, \cdot]_t$, then $K_t = P_t - Q_t$ is the projection on $\mathscr{S}_t$ orthogonal with respect to $[\cdot, \cdot]_t$ and $\mathbf{z} = K_t\mathbf{x} = \mathbf{x} - Q_t\mathbf{x}$.

If

$$\boldsymbol{\lambda}_t = K_t \boldsymbol{\mu}_t$$

and

$$\hat{\boldsymbol{\lambda}}_t = K_t \hat{\boldsymbol{\mu}}_t ,$$

then $\mathbf{m}_t = \mathbf{m}^*(\boldsymbol{\lambda}_t)$ and $\hat{\mathbf{m}}_t = \mathbf{m}^*(\hat{\boldsymbol{\lambda}}_t)$. If (2.1) holds, then $\hat{\gamma}_t = \gamma_t(\hat{\boldsymbol{\lambda}}_t)$, so that the distribution of $\hat{\gamma}_t$ depends on the distribution of $\hat{\boldsymbol{\lambda}}_t$.

Asymptotic results are proven under the following condition:

CONDITION 1. For $t \geqq 0$, $\boldsymbol{\mu}_t \in \mathscr{M}_t$ and constants $b_t \geqq 0$, $B_t \geqq 0$, and $f_t > a_t^{\frac{1}{2}}$ are defined so that

$$\textstyle\sum_{i \in I_t} y_i^2 |m_{it}^*(\mathbf{x}) - m_{it}| \leqq b_t ||\mathbf{x} - \boldsymbol{\lambda}_t||_t ||\mathbf{y}||_t^2 , \qquad \mathbf{x}, \mathbf{y} \in \mathscr{S}_t,\ ||\mathbf{x} - \boldsymbol{\lambda}_t||_t \leqq f_t , \tag{2.11}$$

$$(\textstyle\sum_{i \in I_t} y_i z_i [m_{it}^*(\mathbf{x}) - m_{it}])^2 \leqq B_t ||\mathbf{x} - \boldsymbol{\lambda}_t||_t ||\mathbf{y}||_t^2 ||\mathbf{z}||_t^2 , \qquad \mathbf{x}, \mathbf{y} \in \mathscr{S}_t,\ \mathbf{z} \in \mathscr{N}_t,\ ||\mathbf{x} - \boldsymbol{\lambda}_t||_t \leqq f_t . \tag{2.12}$$

As $t \to \infty$, $a_t(b_t + B_t) \to 0$ and $a_t/f_t^2 \to 0$.

REMARK. Note that under the Poisson sampling model, B_t may be set equal to 0.

This condition is essentially a requirement that individual coordinates y_i of each element $\mathbf{y} \in \mathscr{S}_t$ must be small relative to the norm $||\mathbf{y}||_t$. Note that

$$\frac{m_{it}^*(\mathbf{x}) - m_{it}}{m_{it}} = \exp(x_i - \lambda_i) - 1$$

under the Poisson sampling model and

$$\frac{m_{it}^*(\mathbf{x}) - m_{it}}{m_{it}} = \frac{\exp(x_i - \lambda_i)}{\sum_{j \in J_{kt}} (m_{jt}/N_{kt}) \exp(x_j - \lambda_j)} - 1$$

under the multinomial sampling model. Since Jensen's inequality implies that

$$\begin{aligned} \textstyle\sum_{j \in J_{kt}} (m_{jt}/N_{kt}) \exp(x_j - \lambda_j) &\geqq \exp \textstyle\sum_{j \in J_{kt}} (m_{jt}/N_{kt})(x_j - \lambda_j) \\ &= \exp 0 \\ &= 1 , \end{aligned}$$

it follows that under either model

$$\left| \frac{m_{it}^*(\mathbf{x}) - m_{it}}{m_{it}} \right| \leqq 2 \max_{i \in I_t} (\exp|x_i - \lambda_i| - 1) .$$

Let d_t be the smallest nonnegative number such that

$$|y_i| \leqq d_t||\mathbf{y}||_t\,, \qquad \mathbf{y} \in \mathscr{S}_t\,.$$

Then

$$\sum_{i\in I_t} y_i^2|m_{it}^*(\mathbf{x}) - m_{it}| \leqq 2[\exp(d_t||\mathbf{x} - \boldsymbol{\lambda}_t||_t) - 1]||\mathbf{y}||_t^2\,, \qquad \mathbf{x}, \mathbf{y} \in \mathscr{S}_t$$

and

$$[\sum_{i\in I_t} y_i z_i|m_{it}^*(\mathbf{x}) - m_{it}|]^2 \leqq 4[\exp(d_t||\mathbf{x} - \boldsymbol{\lambda}_t||_t) - 1]^2||\mathbf{y}||_t^2||\mathbf{z}||_t^2\,, \qquad \mathbf{x}, \mathbf{y} \in \mathscr{S}_t,\ \mathbf{z} \in \mathscr{N}_t\,.$$

Thus Condition 1 holds if Condition 2 holds:

CONDITION 2. For $t \geqq 0$, $\boldsymbol{\mu}_t \in \mathscr{M}_t$. As $t \to \infty$, $a_t d_t \to 0$.

The coefficient d_t satisfies the inequality

$$d_t^2 \leqq 1/\min_{i\in I_t} m_{it}\,.$$

Thus Condition 2 is satisfied under the traditional conditions that I_t, $\mathscr{N}_t$, and $\mathscr{M}_t$ are constant for $t \geqq 0$ and

$$\min_{i\in I_t} m_{it} \to \infty \qquad \text{as} \quad t \to \infty\,.$$

On the other hand, if $\mathscr{N}_t \neq \mathscr{M}_t$ for any $t \geqq 0$, then Condition 1 can only hold if the sum

$$N_t = \sum_{i\in I_t} m_{it}$$

of expected cell counts approaches infinity as t approaches infinity. To verify this claim, assume that Condition 1 holds. Let $\mathbf{y} \in \mathscr{S}_t$, $\mathbf{y} \neq 0$. For sufficiently small $\varepsilon > 0$, (2.11) holds with $\mathbf{x} - \boldsymbol{\lambda}_t = \varepsilon\mathbf{y}$. By letting $\varepsilon \to 0$, one finds that

$$\sum_{i\in I_t} |y_i|^3 m_{it} \leqq b_t||\mathbf{y}||_t^3\,. \tag{2.13}$$

By Cramér (1946, page 176)

$$(N_t^{-1}\sum_{i\in I_t} |y_i|^3 m_{it})^{\frac{1}{3}} \geqq (N_t^{-1}\sum_{i\in I_t} y_i^2 m_{it})^{\frac{1}{2}} = N_t^{-\frac{1}{2}}||\mathbf{y}||_t\,.$$

Thus $b_t \geqq N_t^{-\frac{1}{2}}$. Since $a_t \geqq 1$ for $t \geqq 0$, $b_t \to 0$ and $N_t \to \infty$ as $t \to \infty$.

Given Condition 1, the following theorem is proven in Section 4.

THEOREM 1. *Assume that Condition 1 holds. For $t \geqq 0$, let γ_t be a linear functional on $\mathscr{M}_t$ such that (2.1) holds and such that $\gamma_t(\mathbf{x}) \neq 0$ for some $\mathbf{x} \in \mathscr{M}_t$. Then*

$$(\hat{\gamma}_t - \gamma_{t0})/\sigma(\hat{\gamma}_t) \to_{\mathscr{D}} N(0, 1)\,, \tag{2.14}$$

$$\hat{\sigma}(\hat{\gamma}_t)/\sigma(\hat{\gamma}_t) \to_P 1\,, \tag{2.15}$$

and

$$P\{\hat{\gamma}_t - Z_{\alpha/2}\hat{\sigma}(\hat{\gamma}_t) \leqq \gamma_{t0} \leqq \hat{\gamma}_t + Z_{\alpha/2}\hat{\sigma}(\hat{\gamma}_t)\} \to 1 - \alpha\,, \qquad 0 < \alpha < 1\,, \tag{2.16}$$

where $\to_{\mathscr{D}}$ denotes convergence in distribution and $\to_P$ denotes convergence in probability.

Under Condition 2, the consistency property of the following theorem applies.

THEOREM 2. *Assume that Condition* 2 *holds. For* $t \geqq 0$, *let* $\mathscr{C}_t$ *consist of all* $\mathbf{c} \in R^{I_t}$ *such that*

$$\sum_{i \in I_t} |c_i| \leqq 1$$

and

$$\langle \mathbf{c}, \mathbf{x} \rangle_t = 0 , \qquad \mathbf{x} \in \mathscr{N}_t .$$

Then

$$F_t = \sup\{|\langle \mathbf{c}, \hat{\boldsymbol{\mu}}_t \rangle_t - \langle \mathbf{c}, \boldsymbol{\mu}_t \rangle_t| : \mathbf{c} \in \mathscr{C}_t\} \rightarrow_P 0 .$$

Conditions 1 or 2 hold under a wide variety of circumstances. The following examples provide some indication of the range of applications.

EXAMPLE 1. *Constant models with selected increasing cell means.* In traditional treatments of contingency tables, asymptotic theory is developed under the assumption that for $t \geqq 0$, $I_t = I_0$, and $\boldsymbol{\mu}_t \in \mathscr{M}_t = \mathscr{M}_0$, and $\mathscr{N}_t = \mathscr{N}_0$. It is also assumed that

$$N_t = \sum_{i \in I_t} m_{it} \rightarrow \infty$$

and

$$N_t^{-1} m_{it} \rightarrow c_i > 0 , \qquad i \in I_0 .$$

Under these assumptions, Haberman (1974, pages 80–81) derives formulas equivalent to (2.14), (2.15) and (2.16) for the case in which (2.1) holds, $\gamma_0 \neq 0$, and $\gamma_t = \gamma_0$ is a linear functional on $\mathscr{M}_0$ for $t \geqq 0$. Haberman's (1974) results also imply the conclusion of Theorem 2.

A simple generalization of results in Haberman (1974) is available. Assume that for $t \geqq 0$, $I_t = I_0$, $\boldsymbol{\mu}_t \in \mathscr{M}_t = \mathscr{M}_0$, $\mathscr{N}_t = \mathscr{N}_0$, and γ_t is a linear functional on $\mathscr{M}_t$, (2.1) holds, and $\gamma_t \neq 0$. Let $J \subset I_0$ be a given set such that if $\mathbf{x} \in \mathscr{M}_0$ and $x_i = 0$, $i \in J$, then $\mathbf{x} = \mathbf{0}$. Let $m_{it} \rightarrow \infty$, $i \in J$. Then (2.13), (2.14) and (2.15) hold and $F_t \rightarrow_P 0$.

To verify this claim, note first that $a_t = a_0$ for $t \geqq 0$. Thus it suffices to prove that $d_t \rightarrow 0$. To do so, note that for some $\mathbf{x}_t \in \mathscr{S}_t$ and $i(t) \in I_0$,

$$\sum_{i \in I_0} m_{it} x_{it}^2 = 1$$

and

$$d_t = x_{i(t)t} .$$

For each $i \in J$,

$$|x_{it}| \leqq m_{it}^{-\frac{1}{2}} .$$

Let T be the linear mapping from $\mathscr{M}_0$ to R^J, defined so that $\mathbf{y} = T(\mathbf{x})$ if $\mathbf{x} \in \mathscr{M}_0$ and $y_i = x_i$ for all $i \in J$. By assumption, T has a kernel consisting of the zero element $\mathbf{0}$ of $\mathscr{M}_0$. Let $T\mathscr{M}_0$ be the image of $\mathscr{M}_0$ under T. Then for some linear transformation U from $T\mathscr{M}_0$ to $\mathscr{M}_0$,

$$UT\mathbf{x} = \mathbf{x} , \qquad \mathbf{x} \in \mathscr{M}_0 .$$

Since $m_{it} \rightarrow \infty$ for $i \in J$, $T\mathbf{x}_t \rightarrow \mathbf{0}$ and $\mathbf{x}_t = UT\mathbf{x}_t \rightarrow 0$. Thus $d_t \rightarrow 0$.

EXAMPLE 2. *Multinomial response models.* Following Haberman (1974, pages 352–373), consider tables $\mathbf{n}_t$ such that for $t \geqq 0$,

$$I_t = J \times K_t .$$

Let each table $\mathbf{n}_{kt} = \{n_{jkt} : j \in J\}$, $k \in K_t$, be an independent multinomial vector with sample size $N_{kt} > 0$ and probabilities $\{p_{jkt} : j \in J\}$. Define the parameter $\boldsymbol{\theta}_t = \{\theta_{jkt}\}$ by the equations

$$p_{jkt} = \exp\theta_{jkt} / \textstyle\sum_{j' \in J} \exp\theta_{j'kt}, \qquad j \in J,\ k \in K_t \tag{2.17}$$

and

$$\textstyle\sum_{j \in J} \theta_{jkt} = 0, \qquad k \in K_t. \tag{2.18}$$

Assume that associated with any $k \in K_t$ is a known concomitant vector $\mathbf{X}_k \in R^H$ such that for some unknown $\boldsymbol{\beta}_t \in R^{H \times J}$,

$$\theta_{jkt} = \textstyle\sum_{h \in H} X_{hk} \beta_{hjt}, \qquad j \in J,\ k \in K_t. \tag{2.19}$$

Let $\boldsymbol{\beta}_t \in \mathscr{B}$, a linear manifold in $R^{H \times J}$ such that

$$\textstyle\sum_{j \in J} w_{hj} = 0, \qquad h \in H,\ \mathbf{w} \in \mathscr{B}. \tag{2.20}$$

Since the constraint (2.20) implies the constraint (2.18), the model is defined by (2.17) and (2.19). To simplify analysis, assume that $\boldsymbol{\beta}_t$ is uniquely determined by $\boldsymbol{\theta}_t$ for each $t \geqq 0$. In the resulting log-linear model for $\mathbf{n}_t$, $\mathscr{M}_t$ consists of vectors $\mathbf{y} = \{y_{jk} : j \in J,\ k \in K_t\}$ such that for some z_k, $k \in K_t$, and $\mathbf{w} \in \mathscr{B}$,

$$y_{jk} = z_k + \textstyle\sum_{h \in H} X_{hk} w_{hj}, \qquad j \in J,\ k \in K_t.$$

The linear manifold $\mathscr{N}_t$ consists of vectors $\mathbf{y} = \{y_{jk} : j \in J,\ k \in K_t\}$ such that

$$y_{jk} = y_{j'k}, \qquad j, j' \in J,\ k \in K_t,$$

and $\mathscr{S}_t$ consists of vectors $\mathbf{y} = \{y_{jk} : j \in J,\ k \in K_t\}$ such that for some $\mathbf{w} \in \mathscr{B}$,

$$y_{jk} = \textstyle\sum_{h \in H} X_{hk}(w_{hj} - \bar{w}_{hkt}), \qquad j \in J,\ k \in K_t,$$

where

$$\bar{w}_{hkt} = \textstyle\sum_{j \in J} p_{jkt} w_{hj}, \qquad h \in H,\ k \in K_t.$$

The dimension a_t of $\mathscr{S}_t$ is equal to $\dim \mathscr{B}$ for all $t \geqq 0$.

For finite sets A and B and for linear transformations X on R^A and Y on R^B, let

$$\begin{aligned} \|\mathbf{x}\|_A^2 &= \textstyle\sum_{a \in A} x_a^2, & \mathbf{x} &\in R^A, \\ (\mathbf{x}, \mathbf{y})_A &= \textstyle\sum_{a \in A} x_a y_a, & \mathbf{x}, \mathbf{y} &\in R^A, \\ [\mathbf{x} \otimes \mathbf{y}]\mathbf{z} &= \mathbf{x}(\mathbf{y}, \mathbf{z})_B, & \mathbf{x} \in R^A,\ \mathbf{y}, \mathbf{z} &\in R^B, \\ \mathbf{x} * \mathbf{y} &= \{x_a y_b : a \in A,\ b \in B\} & \mathbf{x} \in R^A,\ \mathbf{y} &\in R^B, \end{aligned}$$

and let $X * Y$ be the unique linear transformation on $R^{A \times B}$ such that

$$(X * Y)(\mathbf{x} * \mathbf{y}) = (X\mathbf{x}) * (Y\mathbf{y}), \qquad \mathbf{x} \in R^A,\ \mathbf{y} \in R^B.$$

Let

$$\begin{aligned} g_t &= \max_{k \in K_t} \|\mathbf{X}_k\|_H, \\ W_{kt}\mathbf{x} &= N_k\{p_{jkt}x_j - p_{jkt} \textstyle\sum_{j' \in J} p_{j'kt} x_{j'} : j \in J\}, \qquad \mathbf{x} \in R^J, \\ V_t &= \textstyle\sum_{k \in K_t} [\mathbf{X}_k \otimes \mathbf{X}_k] * W_{kt}, \end{aligned}$$

and

$$h_t = \min\{(\mathbf{w}, V_t\mathbf{w})_{H\times J} : \mathbf{w} \in \mathscr{B}, ||\mathbf{w}||_{H\times J} = 1\} .$$

Since $\boldsymbol{\beta}_t$ is uniquely determined by $\boldsymbol{\theta}_t$, $h_t > 0$. If $g_t/h_t^{\frac{1}{2}} \to 0$ as $t \to \infty$, then $a_t d_t \to 0$ and (2.13), (2.14) and (2.15) hold. Since $F_t \to_P 0$, it is easily shown that

$$\max_{k\in K_t} \max_{j\in J} |\hat{\theta}_{jkt} - \theta_{jkt}| \to_P 0 .$$

To verify the claim that $a_t d_t \to 0$, note that if $\mathbf{x} \in \mathscr{S}_t$ and $||\mathbf{x}||_t = 1$, then for some $\mathbf{w} \in \mathscr{B}$,

$$x_{jk} = \sum_{h\in H} X_{hk}(w_{hj} - \sum_{j'\in J} p_{j'kt} w_{hj'})$$

and

$$(\mathbf{w}, V_t\mathbf{w})_{H\times J} = 1 .$$

Then $||\mathbf{w}||_{H\times J} \leqq h_t^{-\frac{1}{2}}$ and for each $j \in J$ and $k \in K_t$, $|x_{jk}| \leqq 2g_t/h_t^{\frac{1}{2}}$. Thus $d_t \leqq 2g_t/h_t^{\frac{1}{2}}$. Since $a_t = \dim \mathscr{B}$, $a_t d_t \to 0$.

The condition $g_t/h_t^{\frac{1}{2}} \to 0$ is quite weak. For example, it holds if $\{\boldsymbol{\beta}_t : t \geqq 0\}$ is a bounded sequence,

$$\limsup_{t\to\infty} g_t < \infty ,$$

$$K_t \subset K_{t+1} ,$$

the $\mathbf{X}_k$, $k \in K_{t'}$, span R^H for some $t' \geqq 0$, and

$$N_t = \sum_{k\in K_t} N_{kt} \to \infty .$$

The condition $g_t/h_t^{\frac{1}{2}} \to 0$ also holds under the assumptions in Haberman (1974).

EXAMPLE 3. *Longitudinal observations.* For $t \geqq 0$, let $N_t > 0$ subjects be observed at times 1 to $v_t \geqq 2$. Assume that at any time, a subject can be in a state s contained in a finite nonempty set S with q elements. Let s_{uj} denote the state subject j is in at time u, where $1 \leqq j \leqq N_t$ and $1 \leqq u \leqq v_t$. Assume that for some $\pi_s > 0$, $s \in S$, and $\pi_{ss'u} > 0$, $s, s' \in S$, $1 \leqq u \leqq v_t - 1$,

$$P\{s_{1j} = s\} = \pi_s , \qquad s \in S,\ 1 \leqq j \leqq N , \tag{2.21}$$

and

$$P\{s_{(u+1)j} = x(u+1) \mid s_{vj} = x(v),\ 1 \leqq v \leqq u\} = \pi_{x(u)x(u+1)u} , \tag{2.22}$$
$$x(v) \in S,\ 1 \leqq v \leqq u+1,\ 1 \leqq u \leqq v_t - 1,\ 1 \leqq j \leqq N_t .$$

In other words, assume that the subject's state at time $u + 1$, $1 \leqq u \leqq v_t - 1$, is only influenced by the subject's state at time u.

Let $I_t = S^{v_t}$, and let n_{it}, $i \in I_t$, be the number of subjects j, $1 \leqq j \leqq N_t$, such that $s_{uj} = i_u$ for $1 \leqq u \leqq v_t$. Given (2.21) and (2.22),

$$\mu_{it} = \log(N_t \prod_{u=1}^{v_t-1} \pi_{i_u i_{u+1} u}) .$$

Thus equations (2.21) and (2.22) imply that $\boldsymbol{\mu} \in \mathscr{L}_t$, where $\mathbf{x} \in \mathscr{L}_t$ if for some $\mathbf{y}_u \in R^{S\times S}$, $1 \leqq u \leqq v_t - 1$,

$$x_i = \sum_{u=1}^{v_t-1} y_{i_u i_{u+1} u} , \qquad i \in I_t .$$

Let $\mathscr{M}_t$ be a linear manifold contained in $\mathscr{L}_t$ such that $\mathscr{N}_t = \operatorname{span}\{\mathbf{e}_t\} \subset \mathscr{M}_t$. Consider the log-linear model $\boldsymbol{\mu}_t \in \mathscr{M}_t$.

The condition $a_t d_t^2 \to 0$ holds if

$$a_t^2 \max_{i \in I_t} \sum_{u=1}^{v_t - 1} (1/M_{i_u i_{u+1} ut}) \to 0 \,, \tag{2.23}$$

where for $1 \leqq u \leqq v_t - 1$, $s, s' \in S$, and $t \geqq 0$, $M_{ss'ut}$ is the expected number of subjects j, $1 \leqq j \leqq N_t$, such that $s_{uj} = s$ and $s_{(u+1)j} = s'$.

To verify this claim, note that

$$d_t \leqq \max \sup_{i \in I_t} \{|x_i| : \mathbf{x} \in \mathscr{L}_t, \ ||\mathbf{x}||_t = 1\} \,.$$

Let G_t be the projection on $\mathscr{L}_t$ orthogonal with respect to $[\cdot, \cdot]_t$ and let $\boldsymbol{\delta}_t(i) \in R^{I_t}$ be defined for $i \in I_t$ so that

$$\begin{aligned} \delta_{jt}(i) &= 1 \,, \quad j = i \,, \\ &= 0 \,, \quad j \neq i \,. \end{aligned}$$

Then by Haberman (1974) or Rao (1973, page 60),

$$|x_i| = |\langle \mathbf{x}, G_t D_t^{-1} \boldsymbol{\delta}_t(i) \rangle_t| \,, \qquad \mathbf{x} \in \mathscr{L}_t, \ i \in I_t \,,$$

and

$$d_t \leqq \max_{i \in I_t} \langle \boldsymbol{\delta}_t(i), G_t D_t^{-1} \boldsymbol{\delta}_t(i) \rangle_t^{\frac{1}{2}} \,.$$

By Haberman (1974, Chapter 5), $\mathscr{L}_t$ is decomposable and

$$\langle \boldsymbol{\delta}_t(i), G_t D_t^{-1} \boldsymbol{\delta}_t(i) \rangle_t = \sum_{u=1}^{v_t - 1} (1/M_{i_u i_{u+1} ut}) - \sum_{u=1}^{v_t} (1/L_{i_u ut}) \,, \qquad i \in I_t \,,$$

where for $s \in S$, $1 \leqq u \leqq v_t$, L_{sut} is the expected number of subjects j, $1 \leqq j \leqq N_t$, such that $s_{uj} = s$. Therefore, (2.23) implies that $a_t d_t \to 0$.

As one indication of the implications of (2.23), note that if

$$\liminf_{u \to \infty} N_t^{-1} L_{sut} > 0 \,, \qquad s \in S \,,$$

and $a_t^2 v_t / N_t \to 0$, then $a_t d_t \to 0$. Stronger results are available for specially selected $\mathscr{M}_t$, but these general results are sufficient to suggest the applicability of asymptotic theory in longitudinal studies.

2.1. *Comparison of $\hat{\boldsymbol{\mu}}_t$ and $\hat{\boldsymbol{\mu}}_t'$.* To compare the asymptotic distributions of $\hat{\boldsymbol{\mu}}_t$ and $\hat{\boldsymbol{\mu}}_t'$, linear functionals γ_t on $\mathscr{M}_t'$ are considered such that (2.1) holds. Let $\hat{\gamma}_t' = \gamma_t(\hat{\boldsymbol{\mu}}_t')$,

$$\sigma^2(\hat{\gamma}_t') = \sup |\gamma_t(\mathbf{x})|^2 : \mathbf{x} \in \mathscr{M}_t', \ \sum_{i \in I_t} x_i^2 m_{it} = 1\} \,, \tag{2.24}$$

and

$$\hat{\sigma}^2(\hat{\gamma}_t') = \sup \{|\gamma_t(\mathbf{x})|^2 : \mathbf{x} \in \mathscr{M}_t', \ \sum_{i \in I_t} x_i^2 \hat{m}_{it}' = 1\} \,. \tag{2.25}$$

Assume that $\gamma_t(\mathbf{x}) \neq 0$ for some $\mathbf{x} \in \mathscr{M}_t$. To provide regularity conditions for results of this section, let

$$\mathscr{S}_t' = \{\mathbf{y} \in \mathscr{M}_t' : [\mathbf{x}, \mathbf{y}]_t = 0, \mathbf{x} \in \mathscr{N}_t\} \tag{2.26}$$

and

(2.27) $$a_t' = \dim \mathscr{S}_t' = \dim \mathscr{M}_t' - \dim \mathscr{N}_t .$$

Results of this subsection require the following analogue of Condition 1:

CONDITION 1′. For $t \geqq 0$, $\boldsymbol{\mu}_t \in \mathscr{M}_t$ and constants $b_t' \geqq 0$, $B_t' \geqq 0$, and $f_t' > (a_t')^{\frac{1}{2}}$ are defined such that

(2.28) $$\sum_{i \in I_t} y_i^2 |m_{it}^*(\mathbf{x}) - m_{it}| \leqq b_t' ||\mathbf{x} - \boldsymbol{\lambda}_t||_t ||\mathbf{y}||_t^2 , \qquad \mathbf{x}, \mathbf{y} \in \mathscr{S}_t', \ ||\mathbf{y} - \boldsymbol{\lambda}_t||_t \leqq f_t' ,$$

and

(2.29) $$(\sum_{i \in I_t} y_i z_i [m_{it}^*(\mathbf{x}) - m_{it}])^2 \leqq B_t' ||\mathbf{x} - \boldsymbol{\lambda}_t||_t ||\mathbf{y}||_t^2 ||\mathbf{z}||_t^2 , \qquad \mathbf{x}, \mathbf{y} \in \mathscr{S}_t', \ \mathbf{z} \in \mathscr{N}_t, \ ||\mathbf{x} - \boldsymbol{\lambda}_t|| \leqq f_t' .$$

As $t \to \infty$, $a_t'(b_t' + B_t') \to 0$ and $a_t'/(f_t')^2 \to 0$.

Let $d_t' \geqq 0$ be the smallest nonnegative number such that

(2.30) $$|y_i| \leqq d_t' ||\mathbf{y}||_t , \qquad \mathbf{y} \in \mathscr{S}_t' .$$

Then Condition 1′ follows from Condition 2′:

CONDITION 2′. For $t \geqq 0$, $\boldsymbol{\mu}_t \in \mathscr{M}_t$. As $t \to \infty$, $a_t' d_t' \to 0$.

Under Condition 1′, Theorem 1 implies that as $t \to \infty$,

(2.31) $$(\hat{\gamma}_t' - \gamma_{t0})/\sigma(\hat{\gamma}_t') \to_{\mathscr{D}} N(0, 1) ,$$

(2.32) $$\hat{\sigma}(\hat{\gamma}_t')/\sigma(\hat{\gamma}_t') \to_P 1 ,$$

and

(2.33) $$P\{\hat{\gamma}_t' - Z_{\alpha/2}\hat{\sigma}(\hat{\gamma}_t') \leqq \gamma_{t0} \leqq \hat{\gamma}_t' + Z_{\alpha/2}\hat{\sigma}(\hat{\gamma}_t')\} \to 1 - \alpha , \qquad 0 < \alpha < 1 .$$

Since Condition 1′ implies Condition 1, (2.14), (2.15) and (2.16) also hold.

The definitions of $\sigma(\hat{\gamma}_t)$ and $\sigma(\hat{\gamma}_t')$ imply that $\sigma(\hat{\gamma}_t) \leqq \sigma(\hat{\gamma}_t')$, so that the asymptotic variance of $\hat{\gamma}_t$ does not exceed the asymptotic variance of $\hat{\gamma}_t'$. If $\mathscr{M}_t \neq \mathscr{M}_t'$, then $0 < \sigma(\hat{\gamma}_t) < \sigma(\hat{\gamma}_t')$ for some linear functional γ_t on $\mathscr{M}_t'$. Consequently, if $\boldsymbol{\mu}_t \in \mathscr{M}_t$ and $\mathscr{M}_t \neq \mathscr{M}_t'$, then $\hat{\boldsymbol{\mu}}_t$ is a more efficient estimate of $\boldsymbol{\mu}$ than is $\hat{\boldsymbol{\mu}}_t'$. If $\sigma(\hat{\gamma}_t)/\sigma(\hat{\gamma}_t') \to \rho < 1$, then

$$\hat{\sigma}(\hat{\gamma}_t)/\hat{\sigma}(\hat{\gamma}_t') \to_P \rho ,$$

so that for any α, $0 < \alpha < 1$, the probability approaches 1 that the approximate level-$(1 - \alpha)$ confidence interval

$$\hat{\gamma}_t - Z_{\alpha/2}\hat{\sigma}(\hat{\gamma}_t) \leqq \gamma_{t0} \leqq \hat{\gamma}_t + Z_{\alpha/2}\hat{\sigma}(\hat{\gamma}_t)$$

for γ_{t0} is less wide than the approximate level-$(1 - \alpha)$ confidence interval

$$\hat{\gamma}_t' - Z_{\alpha/2}\hat{\sigma}(\hat{\gamma}_t') \leqq \gamma_{t0} \leqq \hat{\gamma}_t' + Z_{\alpha/2}\hat{\sigma}(\hat{\gamma}_t')$$

for γ_{t0}.

One more subtle feature of the relationship between $\hat{\boldsymbol{\mu}}_t$ and $\hat{\boldsymbol{\mu}}_t'$ is given in Theorem 3, which is proven in Section 3.

THEOREM 3. *Assume that Condition* 1′ *is satisfied, and assume that for* $t \geqq 0$, γ_t *is a linear functional on* $\mathscr{M}_t'$ *such that* (2.1) *holds and* $\gamma_t(\mathbf{x}) \neq 0$ *for some* $\mathbf{x} \in \mathscr{M}_t$. *Let*

$$W_t = (\hat{\gamma}_t - \gamma_{t0})/\sigma(\hat{\gamma}_t)$$

and

$$W_t' = (\hat{\gamma}_t' - \gamma_{t0})/\sigma(\hat{\gamma}_t') .$$

If $\sigma(\hat{\gamma}_t)/\sigma(\hat{\gamma}_t') \to \rho$, *then* (W_t, W_t') *converges in distribution to a multivariate normal distribution with mean* $\mathbf{0}$ *and covariance matrix*

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

3. Asymptotic properties of X_t^2 and L_t^2. The chi-square statistics X_t^2 and L_t^2 retain their customary asymptotic distributions under Condition 1′ if $a_t' - a_t \to a$, where $0 < a < \infty$. The following theorem is proven is Section 4:

THEOREM 4. *Assume that Condition* 1′ *is satisfied and* $a_t' - a_t \to a > 0$. *Then*

$$X_t^2 \to_{\mathscr{D}} \chi_a^2 , \tag{3.1}$$

$$L_t^2 \to_{\mathscr{D}} \chi_a^2 , \tag{3.2}$$

and

$$X_t^2 - L_t^2 \to_P 0 . \tag{3.3}$$

The following examples illustrate application of these results.

EXAMPLE 1. *Constant models with selected increasing cell means* (*continued*). In addition to the previous assumptions in this example, assume that $\mathscr{M}_t' = \mathscr{M}_0'$ for $t \geqq 0$ and $\mathscr{M}_0 \neq \mathscr{M}_0'$, and assume that if $\mathbf{x} \in \mathscr{M}_0'$ and $x_i = 0$ for $i \in J$, then $\mathbf{x} = \mathbf{0}$. Then (3.1), (3.2) and (3.3) hold with $a = \dim \mathscr{M}_0' - \dim \mathscr{M}_0$.

EXAMPLE 2. *Multinomial response models* (*continued*). Let $\mathscr{B}'$ be a linear manifold in $R^{H \times J}$ such that $\mathscr{B} \subset \mathscr{B}'$, $\mathscr{B} \neq \mathscr{B}'$, and

$$\textstyle\sum_{j \in J} w_{hj} = 0 , \qquad h \in H,\ \mathbf{w} \in \mathscr{B}' .$$

Consider the model in which (2.16) and (2.18) are assumed to hold for some $\boldsymbol{\beta}_t$ in $\mathscr{B}'$. The corresponding linear manifold $\mathscr{M}_t'$ consists of $\mathbf{y} \in R^{J \times K_t}$ such that for some z_k, $k \in K_t$, and $\mathbf{w} \in \mathscr{B}'$,

$$y_{jk} = z_k + \textstyle\sum_{h \in H} X_{hk} \beta_{hj} , \qquad j \in J,\ k \in K_t .$$

Let

$$h_t' = \min \{(\mathbf{w}, V_t \mathbf{w})_{H \times J} : \mathbf{w} \in \mathscr{B}', ||\mathbf{w}||_{H \times J} = 1\} .$$

For each $t \geqq 0$, assume that $\boldsymbol{\beta}_t \in \mathscr{B}'$ is uniquely determined by $\boldsymbol{\theta}_t$, so that $h_t' > 0$. Assume that $g_t/(h_t')^{\frac{1}{2}} \to 0$. Then (3.1), (3.2) and (3.3) hold with $a = \dim \mathscr{B}' - \dim \mathscr{B}$.

EXAMPLE 3. *Longitudinal observations* (*continued*). Assume that $\mathscr{M}_t \subset \mathscr{M}_t' \subset \mathscr{L}_t$ and $\dim \mathscr{M}_t' - \dim \mathscr{M}_t \to a > 0$. Assume that $\boldsymbol{\mu}_t \in \mathscr{M}_t$ for $t \geqq 0$ and assume that

$$(a_t')^2 \max_{i \in I_t} \sum_{u=1}^{v_t-1} (1/M_{i_u i_{u+1} ut}) \to 0 .$$

Then (3.1), (3.2) and (3.3) hold.

4. Proof of Theorems 1 and 2. The proofs of Theorems 1 and 2 depend on results of Kantorovich and Akilov (1964, pages 697–700) concerning successive approximations to fixed points. These results permit $\hat{\gamma}_t$ to be approximated by $\gamma_{t0} + \gamma_t(K_t D_t^{-1}(\mathbf{n}_t - \mathbf{m}_t))$, a linear function of the observations $\mathbf{n}_t$. This linear function will be shown to have the asymptotic normality properties required for a proof of Theorem 1 as well as the consistency properties required in Theorem 2.

The essential steps in the proof of Theorem 1 are provided by the following lemmas.

LEMMA 1. *Let*

$$H_t(\mathbf{x}) = \mathbf{x} + K_t D_t^{-1}[\mathbf{n}_t - \mathbf{m}_t^*(\mathbf{x})] , \qquad \mathbf{x} \in \mathscr{S}_t . \tag{4.1}$$

If $\mathbf{z}$ *is a fixed point of* H_t, *then* $\hat{\boldsymbol{\mu}}_t$ *exists*, $\hat{\boldsymbol{\lambda}}_t = \mathbf{z}$, *and* $\hat{\mathbf{m}}_t = \mathbf{m}_t^*(\mathbf{z})$. *Conversely, if* $\hat{\boldsymbol{\mu}}_t$ *exists, then* $\hat{\boldsymbol{\lambda}}_t$ *is a fixed point of* H_t.

PROOF. Let $\mathbf{z}$ be a fixed point of H_t. Then

$$\begin{aligned} \langle \mathbf{n}_t - \mathbf{m}_t^*(\mathbf{z}), \mathbf{x} \rangle_t &= [D_t^{-1}[\mathbf{n}_t - \mathbf{m}_t^*(\mathbf{z})], \mathbf{x}]_t \\ &= [K_t D_t^{-1}[\mathbf{n}_t - \mathbf{m}_t^*(\mathbf{z})], \mathbf{x}]_t \\ &= 0 , \qquad \mathbf{x} \in \mathscr{M}_t . \end{aligned} \tag{4.2}$$

Since $\{\boldsymbol{\mu}_t^*(\mathbf{z})\} \in \mathscr{M}_t$, (1.1) implies that $\hat{\mathbf{m}}_t = \mathbf{m}_t^*(\mathbf{z})$. Thus $\hat{\boldsymbol{\lambda}}_t = \mathbf{z}$.

On the other hand, assume $\hat{\boldsymbol{\mu}}_t$ exists. If $\mathbf{z} = \hat{\boldsymbol{\lambda}}_t$, then (4.2) holds. Given (4.2), it follows that $H_t(\mathbf{z}) = \mathbf{z}$. $\square$

LEMMA 2. *Let* $Z_t = ||K_t D_t^{-1}(\mathbf{n}_t - \mathbf{m}_t)||_t^2$. *Then*

$$E(Z_t) = a_t . \tag{4.3}$$

PROOF. Let $\mathbf{c}_{jt}$, $1 \leqq j \leqq a_t$, be an orthonormal basis of $\mathscr{S}_t$ with respect to $[\cdot, \cdot]_t$. Then

$$Z_t = \sum_{j=1}^{a_t} \langle \mathbf{c}_{jt}, \mathbf{n}_t - \mathbf{m}_t \rangle_t^2 . \tag{4.4}$$

As noted in Haberman (1974, pages 7 and 12), $\mathbf{n}_t$ has covariance operator $D_t(E_t - Q_t)$ with repsect to $\langle \cdot, \cdot \rangle_t$, where E_t is the identity operator on R^{I_t}. Since

$$E\langle \mathbf{c}_{jt}, \mathbf{n}_t - \mathbf{m}_t \rangle_t^2 = \langle \mathbf{c}_{jt}, D_t(E_t - Q_t)\mathbf{c}_{jt} \rangle_t = ||\mathbf{c}_{jt}||_t^2 = 1 , \tag{4.5}$$

(4.3) follows. $\square$

LEMMA 3. *Let* g *be the Berry–Esseen constant, and let* Φ *be the normal distribution function. For* $t \geqq 0$, *let* γ_t *be a linear functional on* $\mathscr{M}_t$ *such that* (2.1) *holds,* *and assume that* $\boldsymbol{\mu}_t \in \mathscr{M}_t$ *and* $\gamma_t(\mathbf{x}) \neq 0$ *for some* $\mathbf{x} \in \mathscr{M}_t$. *Then*

$$\text{(4.6)} \qquad \begin{aligned} \Delta_t(x) &= |P\{[\gamma_t(H_t(\boldsymbol{\lambda}_t)) - \gamma_t(\boldsymbol{\mu}_t)]/\sigma(\hat{\gamma}_t) \leqq x\} - \Phi(x)| \\ &\leqq g b_t , \qquad x \in R . \end{aligned}$$

If $b_t \to 0$, *then*

$$\text{(4.7)} \qquad [\gamma_t(H_t(\boldsymbol{\lambda}_t)) - \gamma_t(\boldsymbol{\mu}_t)]/\sigma(\hat{\gamma}_t) \to_{\mathscr{D}} N(0, 1) .$$

REMARK. The constant g has the property that if X_k, $1 \leqq k \leqq N$, are independent random variables with mean 0 and if

$$\operatorname{Var}(X_k) = \sigma_k^2 , \qquad 1 \leqq k \leqq N ,$$
$$E(|X_k|^3) = \xi_k , \qquad 1 \leqq k \leqq N ,$$

then

$$|P\{\textstyle\sum_{k=1}^N X_k/(\sum_{k=1}^N \sigma_k^2)^{\frac{1}{2}} \leqq x\} - \Phi(x)| \leqq g(\sum_{k=1}^N \xi_k)/(\sum_{k=1}^N \sigma_k^2)^{\frac{3}{2}} .$$

Various bounds on g are discussed in Feller (1971, page 544).

PROOF. For $t \geqq 0$, there exists $\mathbf{c}_t \in \mathscr{M}_t$ such that

$$\gamma_t(\mathbf{x}) = [\mathbf{c}_t, \mathbf{x}]_t , \qquad \mathbf{x} \in \mathscr{M}_t .$$

Given (2.1), it follows that $\mathbf{c}_t \in \mathscr{S}_t$ and $\sigma(\hat{\gamma}_t) = ||\mathbf{c}_t||_t$. Thus

$$\begin{aligned} [\sigma(\hat{\gamma}_t)]^{-1}[\gamma_t(H_t(\boldsymbol{\lambda}_t)) - \gamma_t(\boldsymbol{\mu}_t)] &= [\sigma(\hat{\gamma}_t)]^{-1}[\gamma_t(H_t(\boldsymbol{\lambda}_t)) - \gamma_t(\boldsymbol{\lambda}_t)] \\ &= ||\mathbf{c}_t||_t^{-1}\langle \mathbf{c}_t, \mathbf{n}_t - \mathbf{m}_t\rangle_t . \end{aligned}$$

Under either Poisson or multinomial sampling,

$$\text{(4.8)} \qquad \Delta_t(x) \leqq g(\textstyle\sum_{i \in I_t} m_{it}|c_{it}|^3)/(\sum_{i \in I_t} m_{it} c_{it}^2)^{\frac{3}{2}} , \qquad x \in R .$$

To prove (4.8), first assume that $\mathbf{n}_t$ is obtained by Poisson sampling. Then for any $v \geqq 1$, $\langle \mathbf{c}_t, \mathbf{n}_t - \mathbf{m}_t\rangle_t$ has the same distribution as

$$Y = \textstyle\sum_{i \in I_t} \sum_{u=1}^{v} c_{it}(X_{iu} - v^{-1}m_{it}) ,$$

where the X_{iu} are independent Poisson random variables with respective means $v^{-1}m_{it}$. Since

$$\operatorname{Var}(X_{iu} - v^{-1}m_{it}) = v^{-1}m_{it}$$

and

$$E|X_{iu} - v^{-1}m_{it}|^3 \leqq E(|X_{iu}|^3) = E(X_{iu}^3) = v^{-1}m_{it} + 3(v^{-1}m_{it})^2 + (v^{-1}m_{it})^3 ,$$

the Berry–Esseen theorem implies that

$$\Delta_t(x) \leqq g[\textstyle\sum_{i \in I_t} m_{it}|c_{it}|^3(1 + 3v^{-1}m_{it} + v^{-2}m_{it}^2)]/(\sum_{i \in I_t} m_{it} c_{it}^2)^{\frac{3}{2}} , \qquad x \in R .$$

Since v can be made arbitrarily large, (4.8) follows.

Now assume that for $1 \leqq k \leqq r_t$, the $\{n_{it} : i \in J_{kt}\}$ are independent multinomial random vectors. Then $\langle \mathbf{c}_t, \mathbf{n}_t - \mathbf{m}_t\rangle_t$ has the same distribution as

$$Y = \textstyle\sum_{k=1}^{r_t} \sum_{u=1}^{N_{kt}} T_{ku} ,$$

where the X_{ku} are independent random variables such that

$$P\{X_{ku} = i\} = p_{it}, \qquad i \in J_{kt},\ 1 \leqq u \leqq N_{kt},\ 1 \leqq k \leqq r_t,$$

$T_{ku} = T(X_{ku})$, and T is a function on I_t such that $T(i) = c_{it}$, $i \in I_t$. Note that

$$\operatorname{Var}(T_{ku}) = \sum_{i \in J_{kt}} p_{it} c_{it}^2$$

and

$$E(|T_{ku}|^3) = \sum_{i \in J_{kt}} p_{it} |c_{it}|^3 .$$

These equations and the Berry–Esseen theorem imply (4.8).

To derive (4.6) from (4.8), note that (2.13) implies that

$$\sum_{i \in I_t} |c_{it}|^3 m_{it} \leqq b_t \left(\sum_{i \in I_t} c_{it}^2 m_{it}\right)^{\frac{3}{2}} .$$

If $b_t \to 0$, $\Delta_t(x) \to 0$ uniformly in x, so that (4.7) holds. □

LEMMA 4. *Let $dH_{t\mathbf{x}}$ be the differential of H_t at $\mathbf{x} \in \mathscr{S}_t$. Assume* (2.11) *holds for some $b_t \geqq 0$ and $f_t > a_t^{\frac{1}{2}}$. Then for $\mathbf{x} \in \mathscr{S}_t$, $||\mathbf{x} - \boldsymbol{\lambda}_t||_t \leqq f_t$,*

$$(4.9) \qquad ||dH_{t\mathbf{x}}(\mathbf{y})||_t \leqq (b_t + B_t)||\mathbf{x} - \boldsymbol{\lambda}_t||_t ||\mathbf{y}||_t , \qquad \mathbf{y} \in \mathscr{S}_t .$$

REMARK. The differential $dH_{t\mathbf{x}}$ of H_t at $\mathbf{x} \in \mathscr{S}_t$ is the unique linear transformation on $\mathscr{S}_t$ such that

$$||\mathbf{y}||_t^{-1} ||H_t(\mathbf{x} + \mathbf{y}) - H_t(\mathbf{x}) - dH_{t\mathbf{x}}(\mathbf{y})||_t \to 0$$

as $||\mathbf{y}||_t \to 0$.

PROOF. Let $D_t^*(\mathbf{x})$ be defined for $\mathbf{x} \in \mathscr{S}_t$ as the linear transformation on R^{I_t} such that

$$D_t^*(\mathbf{x})\mathbf{y} = \{m_{it}^*(\mathbf{x}) y_i : i \in I_t\} , \qquad \mathbf{y} \in R^{I_t} .$$

Let $Q_t^*(\mathbf{x})$ be the projection on $\mathscr{N}_t$ such that

$$\langle \mathbf{z}, D_t^*(\mathbf{x})\mathbf{y} \rangle_t = \langle \mathbf{z}, D_t^*(\mathbf{x}) Q_t^*(\mathbf{x})\mathbf{y} \rangle_t , \qquad \mathbf{z} \in \mathscr{N}_t,\ \mathbf{z} \in R^{I_t} .$$

Arguments similar to those in Haberman (1974, pages 35, 40 and 41) may be used to show that

$$dH_{t\mathbf{x}}(\mathbf{y}) = \mathbf{y} - K_t D_t^{-1} D_t^*(\mathbf{x})[E_t - Q_t^*(\mathbf{x})]\mathbf{y} , \qquad \mathbf{y} \in \mathscr{S}_t .$$

Since $Q_t^*(\mathbf{x})$ has range $\mathscr{N}_t$ and K_t has null space $\mathscr{N}_t$, and since $K_t \mathbf{y} = \mathbf{y}$, $\mathbf{y} \in \mathscr{S}_t$,

$$dH_{t\mathbf{x}}(\mathbf{y}) = -K_t D_t^{-1}[D_t^*(\mathbf{x}) - D_t][E_t - Q_t^*(\mathbf{x})]\mathbf{y} , \qquad \mathbf{y} \in \mathscr{S}_t .$$

Since

$$\langle \mathbf{y}, D_t^*(\mathbf{x})\mathbf{y} \rangle_t = \langle \mathbf{y}, D_t \mathbf{y} \rangle_t , \qquad \mathbf{y} \in \mathscr{N}_t ,$$

it follows from Haberman (1975) that

$$\begin{aligned} Q_t^*(\mathbf{x}) &= Q_t + Q_t D_t^{-1} D_t^*(\mathbf{x})[E_t - Q_t] \\ &= Q_t + Q_t D_t^{-1}[D_t^*(\mathbf{x}) - D_t][E_t - Q_t] . \end{aligned}$$

Thus

$$dH_{t\mathbf{x}}(\mathbf{y}) = -K_t D_t^{-1}[D_t^*(\mathbf{x}) - D_t]\{\mathbf{y} - Q_t D_t^{-1}[D_t^*(\mathbf{x}) - D_t]\mathbf{y}\} , \qquad \mathbf{y} \in \mathscr{S}_t .$$

If $\mathscr{M}_t = \mathscr{N}_t$, then $dH_{t\mathbf{x}}(\mathbf{y}) = \mathbf{0}$ and (4.9) is trivial. If $\mathscr{M}_t \neq \mathscr{N}_t$, then note that

$$||dH_{t\mathbf{x}}(\mathbf{y})||_t = \sup_{\mathbf{z} \in \mathscr{S}_t; \mathbf{z} \neq \mathbf{0}} [\mathbf{z}, dH_{t\mathbf{x}}(\mathbf{y})]_t / ||\mathbf{z}||_t .$$

For $\mathbf{y}, \mathbf{z} \in \mathscr{S}_t$,

$$[\mathbf{z}, dH_{t\mathbf{x}}(\mathbf{y})]_t = \langle \mathbf{z}, [D_t - D_t^*(\mathbf{x})]\mathbf{y} \rangle_t + \langle \mathbf{z}, [D_t^*(\mathbf{x}) - D_t] Q_t D_t^{-1} [D_t^*(\mathbf{x}) - D_t]\mathbf{y} \rangle_t .$$

It is easily verified that

$$[\mathbf{z}, dH_{t\mathbf{x}}(\mathbf{y})]_t = [\mathbf{y}, dH_{t\mathbf{x}}(\mathbf{z})]_t , \qquad \mathbf{y}, \mathbf{z} \in \mathscr{S}_t .$$

Since $dH_{t\mathbf{x}}$ is symmetrical,

$$||dH_{t\mathbf{x}}(\mathbf{y})||_t \leqq g_t ||\mathbf{y}||_t ,$$

where

$$g_t = \sup_{\mathbf{y} \in \mathscr{S}_t; \mathbf{y} \neq \mathbf{0}} |[\mathbf{y}, dH_{t\mathbf{x}}(\mathbf{y})]_t| / ||\mathbf{y}||_t^2 .$$

To complete the proof, note that if $||\mathbf{x} - \boldsymbol{\lambda}_t||_t \leqq f_t$, $\mathbf{x}, \mathbf{y} \in \mathscr{S}_t$, then

$$|\langle \mathbf{y}, [D_t - D_t^*(\mathbf{x})]\mathbf{y} \rangle_t| \leqq \sum y_i^2 |m_{it}^*(\mathbf{x}) - m_{it}| \leqq b_t ||\mathbf{x} - \boldsymbol{\lambda}_t||_t ||\mathbf{y}||_t^2 .$$

Under the Poisson sampling model, Q_t is the zero operator and

$$[\mathbf{y}, dH_{t\mathbf{x}}(\mathbf{y})]_t = \langle \mathbf{y}, [D_t - D_t^*(\mathbf{x})]\mathbf{y} \rangle_t ,$$

so that the proof is complete for this case.

Under the multinomial sampling model, the proof is completed by noting that

$$\begin{aligned} |\langle \mathbf{y}, [D_t^*(\mathbf{x}) - D_t] Q_t D_t^{-1} [D_t^*(\mathbf{x}) - D_t]\mathbf{y} \rangle| &= ||Q_t D_t^{-1} [D_t^*(\mathbf{x}) - D_t]\mathbf{y}||_t^2 \\ &\leqq \sup_{\mathbf{z} \in \mathscr{N}_t; \mathbf{z} \neq \mathbf{0}} \langle \mathbf{z}, [D_t^*(\mathbf{x}) - D_t]\mathbf{y} \rangle_t^2 / ||\mathbf{z}||_t^2 \end{aligned}$$

and

$$\langle \mathbf{z}, [D_t^*(\mathbf{x}) - D_t]\mathbf{y} \rangle_t = \textstyle\sum_{i \in I_t} y_i z_i [m_{it}^*(\mathbf{x}) - m_{it}] . \qquad \square$$

LEMMA 5. *Assume that Condition* 1 *holds. Then*

$$Z_t^{-\frac{1}{2}} ||\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\lambda}_t||_t \to_P 1 \tag{4.10}$$

and

$$||\hat{\boldsymbol{\lambda}}_t - H_t(\boldsymbol{\lambda}_t)||_t \to_P 0 . \tag{4.11}$$

REMARK. To include the case in which $a_t = 0$, 0/0 is defined in (4.10) to be 0.

PROOF. Note that $Z_t = ||H_t(\boldsymbol{\lambda}_t) - \boldsymbol{\lambda}_t||_t^2$. Assume that Condition 1 holds for b_t and f_t. By Lemma 4 and Kantorovich and Akilov (1964, pages 697–700), if $Z_t^{\frac{1}{2}} \leqq \frac{1}{2} f_t$ and $Z_t^{\frac{1}{2}}(b_t + B_t) < \frac{1}{2}$, then H_t has a fixed point $\hat{\boldsymbol{\lambda}}_t$ for which

$$||\boldsymbol{\lambda}_t - H_t(\hat{\boldsymbol{\lambda}}_t)||_t \leqq Z_t^{\frac{1}{2}} \{1 - [1 - 2Z_t^{\frac{1}{2}}(b_t + B_t)]^{\frac{1}{2}}\} / \{1 + [1 - 2Z_t^{\frac{1}{2}}(b_t + B_t)]^{\frac{1}{2}}\} .$$

By the triangle inequality,

$$Z_t^{\frac{1}{2}} - ||\hat{\boldsymbol{\lambda}}_t - H_t(\boldsymbol{\lambda}_t)||_t \leqq ||\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\lambda}_t||_t \leqq Z_t^{\frac{1}{2}} + ||\hat{\boldsymbol{\lambda}}_t - H_t(\boldsymbol{\lambda}_t)||_t .$$

By Lemma 2,

$$P\{Z_t^{\frac{1}{2}} \geqq \tfrac{1}{2}f_t\} \leqq E(Z_t)/(f_t^2/4) = a_t/(f_t^2/4) \to 0\,. \tag{4.12}$$

Since $a_t b_t \to 0$, there exists g_t, $t \geqq 0$, such that $g_t^2 b_t \to 0$ and $a_t/g_t^2 \to 0$. By the same argument used in (4.12),

$$P\{Z_t \geqq g_t\} \to 0\,.$$

Consequently, (4.10) and (4.11) hold. □

These lemmas permit a straightforward proof of Theorem 1. Under the conditions of this theorem, Lemma 5 implies that

$$[\sigma(\hat{\gamma}_t)]^{-1}[\gamma_t(\hat{\boldsymbol{\mu}}_t) - \gamma_t(H_t(\boldsymbol{\lambda}_t))] = [\sigma(\hat{\gamma}_t)]^{-1}[\gamma_t(\hat{\boldsymbol{\lambda}}_t - H_t(\boldsymbol{\lambda}_t))] \to_P 0\,.$$

Lemma 3 then implies (2.14). Lemma 5 and Condition 1 imply (2.15). Equation (2.16) is a simple consequence of (2.14) and (2.15).

To prove Theorem 2, note that under Condition 2,

$$\begin{aligned}\langle \mathbf{c}, \hat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}_t\rangle_t = \langle \mathbf{c}, \hat{\boldsymbol{\lambda}}_t - \boldsymbol{\lambda}_t\rangle_t &\leqq \textstyle\sum_{i\in I_t} |c_i||\hat{\lambda}_{it} - \lambda_{it}| \\ &\leqq d_t(\textstyle\sum_{i\in I_t}|c_i|)||\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\lambda}_t||_t \leqq d_t||\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\lambda}_t||_t\,, \qquad \mathbf{c} \in \mathscr{C}_t\,.\end{aligned}$$

Theorem 2 follows from Lemmas 2 and 5.

5. Proof of Theorem 3. To prove Theorem 3, it suffices to show that for any real η and η',

$$\eta W_t + \eta' W_t' \to_{\mathscr{D}} N(0, \eta^2 + 2\eta\eta'\rho + \eta'^2)\,. \tag{5.1}$$

To prove this claim, let

$$H_t'(\mathbf{x}) = \mathbf{x} + K_t' D_t^{-1}[\mathbf{n}_t - \mathbf{m}_t^*(\mathbf{x})]\,, \qquad \mathbf{x} \in \mathscr{S}_t'\,,$$

where K_t' is the projection on $\mathscr{S}_t'$ orthogonal with respect to $[\cdot, \cdot]_t$. By Lemma 5,

$$\eta W_t + \eta_t W_t' - \eta\gamma_t(H_t(\boldsymbol{\lambda}_t) - \boldsymbol{\lambda}_t)/\sigma(\hat{\gamma}_t) - \eta'\gamma_t(H_t'(\boldsymbol{\lambda}_t) - \boldsymbol{\lambda}_t)/\sigma(\hat{\gamma}_t') \to_P 0\,. \tag{5.2}$$

Let

$$\alpha_t = [\eta/\sigma(\hat{\gamma}_t)]\gamma_t P_t + [\eta'/\sigma(\hat{\gamma}_t')]\gamma_t\,.$$

Note that $\gamma_t P_t(\mathbf{x}) = \gamma_t(P_t\mathbf{x})$ for all $\mathbf{x}$ in R^{I_t}.

For some $\mathbf{c}_t \in \mathscr{M}_t'$

$$\gamma_t(\mathbf{x}) = [\mathbf{c}_t, \mathbf{x}]_t\,, \qquad \mathbf{x} \in \mathscr{M}_t'\,,$$

and

$$\gamma_t(\mathbf{x}) = [P_t\mathbf{c}_t, \mathbf{x}]_t\,, \qquad \mathbf{x} \in \mathscr{M}_t\,.$$

Thus

$$\sigma(\hat{\gamma}_t) = ||P_t\mathbf{c}_t||_t\,,$$
$$\sigma(\hat{\gamma}_t') = ||\mathbf{c}_t||_t\,,$$
$$\alpha_t(\mathbf{x}) = [(\eta/||P_t\mathbf{c}_t||_t)P_t\mathbf{c}_t + (\eta'/||\mathbf{c}_t||_t)\mathbf{c}_t, \mathbf{x}]_t\,, \qquad \mathbf{x} \in \mathscr{M}_t'\,,$$

and

$$\begin{aligned} ||\alpha_t||_t' &= ||(\eta/||P_t\mathbf{c}_t||_t)P_t\mathbf{c}_t + (\eta'/||\mathbf{c}_t||_t)\mathbf{c}_t||_t \\ &= \eta^2 + 2\eta\eta'||P_t\mathbf{c}_t||/||\mathbf{c}_t||_t + \eta'^2 \\ &= \eta^2 + 2\eta\eta'\sigma(\hat{\gamma}_t)/\sigma(\hat{\gamma}_t') + \eta'^2 . \end{aligned} \tag{5.3}$$

Lemma 3 implies that

$$[\eta\gamma_t(H_t(\boldsymbol{\lambda}_t) - \boldsymbol{\lambda}_t)/\sigma(\hat{\gamma}_t) + \eta'\gamma_t(H_t'(\boldsymbol{\lambda}_t) - \boldsymbol{\lambda}_t)/\sigma(\hat{\gamma}_t')]/||\alpha_t||_t' \to_{\mathscr{D}} N(0, 1) . \tag{5.4}$$

Equation (5.1) follows from (5.2), (5.3) and (5.4).

6. Proof of Theorem 4. The proof of Theorem 4 depends on the following lemmas.

LEMMA 6. *Assume that Condition* 1′ *holds and* $a_t' - a_t \to a > 0$. *Then*

$$C_t^2 = ||(K_t' - K_t)D_t^{-1}(\mathbf{n}_t - \mathbf{m}_t)||_t^2 \to_{\mathscr{D}} \chi_a^2 . \tag{6.1}$$

PROOF. Let $\mathscr{W}_t$ be the orthogonal complement of $\mathscr{S}_t$ relative to $\mathscr{S}_t'$ with respect to $[\cdot, \cdot]_t$, so that

$$\mathscr{W}_t = \{\mathbf{x} \in \mathscr{M}_t' : [\mathbf{x}, \mathbf{y}]_t = 0, \mathbf{y} \in \mathscr{M}_t\} .$$

For some $t' \geqq 0$, if $t \geqq t'$, then $\mathscr{W}_t$ has an orthonormal basis $\{\mathbf{c}_{jt} : 1 \leqq j \leqq a\}$ with respect to $[\cdot, \cdot]_t$. Let

$$Z_{jt} = \langle \mathbf{c}_{jt}, \mathbf{n}_t - \mathbf{m}_t \rangle_t , \qquad 1 \leqq j \leqq a .$$

Then

$$||(K_t' - K_t)D_t^{-1}(\mathbf{n}_t - \mathbf{m}_t)||_t^2 = \sum_{j=1}^a Z_{jt}^2 .$$

The lemma follows if it can be shown that $Z_t = \{Z_{jt} : 1 \leqq j \leqq a\}$ converges in distribution to $N(\mathbf{0}, I)$, where I is the identity operator on R^a. To prove this claim, it suffices to show that

$$\sum_{j=1}^a x_j Z_{jt} \to_{\mathscr{D}} N(0, \sum_{j=1}^a x_j^2) , \qquad \mathbf{x} \in R^a .$$

This claim follows from Lemma 3 by noting that

$$\sum_{j=1}^a x_j Z_{jt} = \gamma_t(H_t'(\boldsymbol{\lambda}_t)) - \gamma_t(\boldsymbol{\mu}_t) ,$$

where

$$\gamma_t(\mathbf{y}) = [\sum_{j=1}^a x_j \mathbf{c}_{jt}, \mathbf{y}]_t \qquad \mathbf{y} \in \mathscr{M}_t' ,$$

and

$$||\gamma_t||_t' = ||\sum_{j=1}^a x_j \mathbf{c}_{jt}||_t = (\sum_{j=1}^a x_j^2)^{\frac{1}{2}} . \qquad \square$$

LEMMA 7. *Assume that Condition* 1′ *holds and* $\mathscr{M}_t \neq \mathscr{M}_t'$ *for* $t \geqq 0$. *Then*

$$L_t^2/C_t^2 \to_P 1 . \tag{6.2}$$

PROOF. From Haberman (1973), it follows that the log-likelihood kernel l_t is defined by

$$l_t(\mathbf{x}) = \langle \mathbf{n}_t, \mathbf{x} \rangle_t - \sum_{i \in I_t} \exp x_i , \qquad \mathbf{x} \in R^{I_t} ,$$

under Poisson sampling and

$$l_t(\mathbf{x}) = \langle \mathbf{n}_t, \mathbf{x} \rangle_t - \textstyle\sum_{k=1}^{r_t} N_{kt} \log \sum_{i \in J_{kt}} \exp x_i\,, \qquad \mathbf{x} \in R^{I_t}\,,$$

under multinomial sampling. If $\hat{\boldsymbol{\lambda}}_t' = K_t' \hat{\boldsymbol{\mu}}_t$, then

$$l_t(\hat{\boldsymbol{\lambda}}_t) = \sup_{\mathbf{x} \in \mathscr{M}_t} l_t(\mathbf{x})$$

and

$$l_t(\hat{\boldsymbol{\lambda}}_t') = \sup_{\mathbf{x} \in \mathscr{M}_t'} l_t(\mathbf{x})\,.$$

A Taylor expansion about $\hat{\boldsymbol{\lambda}}_t'$ shows that

$$\begin{aligned} L_t^2 &= 2[l_t(\hat{\boldsymbol{\lambda}}_t') - l_t(\hat{\boldsymbol{\lambda}}_t)] \\ &= -2\langle \mathbf{n}_t - \hat{\mathbf{m}}_t', \hat{\boldsymbol{\lambda}}_t - \hat{\boldsymbol{\lambda}}_t' \rangle_t \\ &\qquad + \langle \hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t, D_t^*(\boldsymbol{\lambda}_t^+)[E_t - Q_t^*(\boldsymbol{\lambda}_t^+)](\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t) \rangle_t \\ &= \langle \hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t, D_t^*(\boldsymbol{\lambda}_t^+)[E_t - Q_t^*(\boldsymbol{\lambda}_t^+)](\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t) \rangle_t\,, \end{aligned} \tag{6.3}$$

where for some z_t, $0 < z_t < 1$,

$$\boldsymbol{\lambda}_t^+ = z_t \hat{\boldsymbol{\lambda}}_t + (1 - z_t)\hat{\boldsymbol{\lambda}}_t'\,.$$

In Lemma 3, it is shown that under Condition 1′, if $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathscr{S}_t'$ and $||\mathbf{x} - \boldsymbol{\lambda}_t||_t \leqq f_t'$, then

$$[\mathbf{z}, dH_{t\mathbf{x}}'(\mathbf{y})]_t = [\mathbf{z}, \mathbf{y}]_t - \langle \mathbf{z}, D_t^*(\mathbf{x})[E_t - Q_t^*(\mathbf{x})]\mathbf{y} \rangle_t$$

has absolute values less than or equal to $(b_t' + B_t')||\mathbf{x} - \boldsymbol{\lambda}||_t ||\mathbf{y}||_t ||\mathbf{z}||_t$. Consequently, Condition 1′ implies that if $||\boldsymbol{\lambda}_t^+ - \boldsymbol{\lambda}_t||_t \leqq f_t'$, then

$$\begin{aligned} |||\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t||_t^2 - L_t^2| &\leqq (b_t' + B_t')||\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t||_t^2 ||\hat{\boldsymbol{\lambda}}_t^+ - \boldsymbol{\lambda}_t||_t \\ &\leqq (b_t' + B_t')||\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t||_t^3\,. \end{aligned}$$

By Lemma 5,

$$||\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t - H_t'(\boldsymbol{\lambda}_t) + H_t(\boldsymbol{\lambda}_t)||_t \to_P 0\,.$$

Since

$$C_t' = ||H_t'(\boldsymbol{\lambda}_t) - H_t(\boldsymbol{\lambda}_t)||_t^2\,,$$

$$||\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t||_t^2 / C_t^2 \to_P 1\,.$$

Since $C_t \leqq Z_t'$ and $Z_t' b_t' \to_P 0$, $L_t^2/C_t^2 \to_P 1$. □

LEMMA 8. *Assume that Condition* 1′ *holds and* $\mathscr{M}_t \neq \mathscr{M}_t'$ *for* $t \geqq 0$. *Then*

$$X_t^2 / C_t^2 \to_P 1\,. \tag{6.4}$$

PROOF. For some z_t, $0 < z_t < 1$, a Taylor expansion shows that

$$\begin{aligned} X_t^2 &= \langle \hat{\mathbf{m}}_t' - \hat{\mathbf{m}}_t, [D_t^*(\hat{\boldsymbol{\lambda}}_t)]^{-1}(\hat{\mathbf{m}}_t' - \hat{\mathbf{m}}_t) \rangle_t \\ &= \langle \hat{\mathbf{m}}_t' - \hat{\mathbf{m}}_t, [D_t^*(\hat{\boldsymbol{\lambda}}_t)]^{-1} Z_t (\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t) \rangle_t\,, \end{aligned}$$

where

$$Z_t = D_t^*(\boldsymbol{\lambda}_t^0)[E_t - Q_t^*(\boldsymbol{\lambda}_t^0)]$$

and

$$\boldsymbol{\lambda}_t^0 = z_t \hat{\boldsymbol{\lambda}}_t + (1 - z_t)\hat{\boldsymbol{\lambda}}_t'\,.$$

A second Taylor expansion shows that

$$\begin{aligned} X_t^2 &= \langle Y_t(\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t), [D_t^*(\hat{\boldsymbol{\lambda}}_t)]^{-1} Z_t(\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t)\rangle_t \\ &= \langle(\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t), Y_t[D_t^*(\hat{\boldsymbol{\lambda}}_t)]^{-1} Z_t(\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t)\rangle_t, \end{aligned}$$

where for some z_t, $0 < z_t < 1$,

$$Y_t = D_t'(\boldsymbol{\lambda}_t^+)[E_t - Q_t^*(\boldsymbol{\lambda}_t^+)]$$

and

$$\boldsymbol{\lambda}_t^+ = y_t\hat{\boldsymbol{\lambda}}_t + (1 - y_t)\hat{\boldsymbol{\lambda}}_t' .$$

As in Lemma 4, these expansions follow from Haberman (1974, pages 35, 40, and 41). Note that $Z_t(\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t)$ is *not* necessarily equal to $\hat{\mathbf{m}}_t' - \hat{\mathbf{m}}$. Thus Y_t and Z_t may differ. If

$$W_t = D_t^*(\hat{\boldsymbol{\lambda}}_t)[E_t - Q_t^*(\hat{\boldsymbol{\lambda}}_t)],$$

then

$$\begin{aligned} X_t^2 = \langle \hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t, Y_t[D_t^*(\hat{\boldsymbol{\lambda}}_t)]^{-1} W_t(\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t)\rangle_t \\ + \langle \hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t, Y_t[D_t^*(\hat{\boldsymbol{\lambda}}_t)]^{-1}(Z_t - W_t)(\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t)\rangle_t . \end{aligned}$$

Using standard properties of projections (see, for example, Rao (1973, page 47)), one finds that

$$\begin{aligned} Y_t[D_t^*(\hat{\boldsymbol{\lambda}}_t)]^{-1} W_t &= D_t^*(\boldsymbol{\lambda}_t^+)[E_t - Q_t^*(\boldsymbol{\lambda}_t^+)][E_t - Q_t^*(\hat{\boldsymbol{\lambda}}_t)] \\ &= D_t^*(\boldsymbol{\lambda}_t^+)[E_t - Q_t^*(\hat{\boldsymbol{\lambda}}_t^+)] \\ &= Y_t . \end{aligned}$$

Proceeding as in the proof of Lemma 4, one finds that

$$\begin{aligned} |\langle \hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t, Y_t[D_t^*(\hat{\boldsymbol{\lambda}}_t)]^{-1} W_t(\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t)\rangle_t - \|\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t\|_t^2| \\ \leqq (b_t' + B_t')\|\boldsymbol{\lambda}_t^+ - \boldsymbol{\lambda}_t\|_t \|\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t\|_t^2 \end{aligned}$$

whenever $\|\boldsymbol{\lambda}_t^+ - \boldsymbol{\lambda}_t\|_t \leqq f_t'$. Since $\|\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t\|_t^2/C_t^2 \to_P 1$ and since

$$\|\boldsymbol{\lambda}_t^+ - \boldsymbol{\lambda}_t\|_t \leqq \max(\|\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t\|_t, \|\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\lambda}_t\|_t),$$

$$\langle \hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t, Y_t[D_t^*(\hat{\boldsymbol{\lambda}}_t)]^{-1} W_t(\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t)\rangle_t / C_t^2 \to_P 1 .$$

The proof is completed by a demonstration that

$$\langle \hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t, Y_t[D_t^*(\hat{\boldsymbol{\lambda}}_t)]^{-1}(Z_t - W_t)(\hat{\boldsymbol{\lambda}}_t' - \hat{\boldsymbol{\lambda}}_t)\rangle_t / C_t^2 \to_P 0 .$$

Such a demonstration is rather tedious, so details are omitted. The principal observation required is that for any $\mathbf{x} \in \mathscr{S}_t$, $\mathbf{x} \neq 0$, $\mathbf{z} \in \mathscr{S}_t$, $\mathbf{z} \neq 0$,

$$\begin{aligned} &\langle \mathbf{x}, Y_t[D_t^*(\hat{\boldsymbol{\lambda}}_t)]^{-1}(Z_t - W_t)\mathbf{x}\rangle_t \\ &\quad \leqq \langle \mathbf{x}, Y_t[D_t^*(\hat{\boldsymbol{\lambda}}_t)]^{-1} Y_t \mathbf{x}\rangle_t^{\frac{1}{2}} \langle \mathbf{x}, (Z_t - W_t)[D_t^*(\hat{\boldsymbol{\lambda}}_t)]^{-1}(Z_t - W_t)\mathbf{x}\rangle_t^{\frac{1}{2}} . \end{aligned}$$

By Rao (1973, page 60), for some $\mathbf{w}, \mathbf{z} \in \mathscr{S}_t$, $\mathbf{w} \neq \mathbf{0}$, $\mathbf{z} \neq \mathbf{0}$,

$$\langle \mathbf{x}, Y_t[D_t^*(\hat{\boldsymbol{\lambda}}_t)]^{-1} Y_t \mathbf{x}\rangle_t = \langle \mathbf{x}, Y_t \mathbf{w}\rangle_t^2 / \langle \mathbf{w}, D_t^*(\hat{\boldsymbol{\lambda}}_t)\mathbf{w}\rangle_t$$

and

$$\langle \mathbf{x}, (Z_t - W_t)[D_t^*(\hat{\boldsymbol{\lambda}}_t)]^{-1}(Z_t - W_t)\mathbf{x}\rangle_t^{\frac{1}{2}} = \langle \mathbf{x}, (Z_t - W_t)\mathbf{z}\rangle_t^2 / \langle \mathbf{z}, D_t^*(\hat{\boldsymbol{\lambda}}_t)\mathbf{z}\rangle_t .$$

If $||\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\lambda}_t||_t \leqq f_t$,

$$|\langle \mathbf{x}, W_t \mathbf{z}\rangle_t - [\mathbf{x}, \mathbf{z}]_t| \leqq (b_t' + B_t')||\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\lambda}_t||_t ||\mathbf{x}||_t ||\mathbf{z}||_t ,$$

$$|\langle \mathbf{w}, D_t^*(\hat{\boldsymbol{\lambda}}_t)\mathbf{w}\rangle_t - ||\mathbf{w}||_t^2| \leqq b_t' ||\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\lambda}_t||_t ,$$

and

$$|\langle \mathbf{z}, D_t^*(\hat{\boldsymbol{\lambda}}_t)\mathbf{z}\rangle_t - ||\mathbf{z}||_t^2| \leqq b_t' ||\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\lambda}_t||_t .$$

If $||\boldsymbol{\lambda}_t^+ - \boldsymbol{\lambda}_t||_t \leqq f_t$,

$$|\langle \mathbf{x}, Y_t \mathbf{w}\rangle_t| \leqq ||\mathbf{x}||_t ||\mathbf{w}||_t [1 + (b_t' + B_t')||\boldsymbol{\lambda}_t^+ - \boldsymbol{\lambda}_t||_t] .$$

If $||\boldsymbol{\lambda}_t^0 - \boldsymbol{\lambda}_t||_t \leqq f_t$,

$$|\langle \mathbf{x}, Z_t \mathbf{w}\rangle_t - [\mathbf{x}, \mathbf{z}]_t| \leqq (b_t' + B_t')||\boldsymbol{\lambda}_t^0 - \boldsymbol{\lambda}_t||_t ||\mathbf{x}||_t ||\mathbf{z}||_t .$$

Given these observations, completion of the proof is not difficult.

REFERENCES

[1] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[2] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

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

[4] HABERMAN, S. J. (1974). *The Analysis of Frequency Data*. Univ. of Chicago Press.

[5] HABERMAN, S. J. (1975). How much do Gauss-Markov and least squares estimates differ? A coordinate-free approach. *Ann. Statist.* **3** 982–990.

[6] HABERMAN, S. J. (1977). Maximum likelihood estimates in exponential response models. *Ann. Statist.* **5** 815–841.

[7] KANTOROVICH, L. V. and AKILOV, G. P. (1964). *Functional Analysis in Normed Spaces* (trans. D. E. Brown). Pergamon Press, Oxford.

[8] RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

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