The Annals of Statistics
1977, Vol. 5, No. 5, 815–841

MAXIMUM LIKELIHOOD ESTIMATES IN EXPONENTIAL RESPONSE MODELS[1]

BY SHELBY J. HABERMAN

University of Chicago

Exponential response models are a generalization of logit models for quantal responses and of regression models for normal data. In an exponential response model, $\{F(\theta): \theta \in \Theta\}$ is an exponential family of distributions with natural parameter θ and natural parameter space $\Theta \subset V$, where V is a finite-dimensional vector space. A finite number of independent observations S_i, $i \in I$, are given, where for $i \in I$, S_i has distribution $F(\theta_i)$. It is assumed that $\boldsymbol{\theta} = \{\theta_i : i \in I\}$ is contained in a linear subspace. Properties of maximum likelihood estimates $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ are explored. Maximum likelihood equations and necessary and sufficient conditions for existence of $\hat{\boldsymbol{\theta}}$ are provided. Asymptotic properties of $\hat{\boldsymbol{\theta}}$ are considered for cases in which the number of elements in I becomes large. Results are illustrated by use of the Rasch model for educational testing.

1. Introduction. Models often arise in statistical practice in which observations have distributions belonging to an exponential family and the corresponding natural parameters satisfy a linear model. Such models, which may be called exponential response models, are considered by Dempster (1971) and Nelder and Wedderburn (1972).

To define an exponential response model, let T be a measurable space with associated σ-algebra $\mathscr{T}$ and σ-finite measure ν. Let V be a finite-dimensional vector space with inner product $(\cdot, \cdot)$, and let Y be a measurable function from T to V. Let Θ consist of $\theta \in V$ such that

$$1/a(\theta) = \int \exp(\theta, Y(s))\, d\nu(s) < \infty ,$$

and for $\theta \in \Theta$, let $F(\theta)$ be the probability distribution on T with density

$$p(\theta, s) = a(\theta) \exp(\theta, Y(s)) , \quad s \in T ,$$

with respect to ν. Let $\mathscr{F}$ be the family of distinct distributions $F(\theta)$, $\theta \in \Theta$. Assume that $\mathscr{F}$ contains more than one element. Let Θ° be the interior of Θ, let C be the convex support of νY^{-1}, and let C° be the interior of C. Note that the family $\mathscr{F}$ may be generated by V, $(\cdot, \cdot)$, and Y chosen so that Θ° and C° are not empty and so that Y has range in C. See Berk (1972) or Barndorff-Nielsen (1973) for justification of this claim. Given this observation, it is assumed in this paper that Θ° and C° are not empty and Y has range in C.

Received November 1975; revised January 1977.

[1] Support for this research was provided in part by NSF Grant No. SOC72-05228 A04 of the Division of Social Sciences, National Science Foundation, and by NSF Grant No. MCS72-04364 A04 of the Division of Mathematical and Computer Sciences, National Science Foundation.

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

Key words and phrases. Exponential family, maximum likelihood estimation, quantal response, logit model, Rasch model, asymptotic theory.

In an exponential response model, independent observations S_i, $i \in I$, are given, where I is a finite and nonempty index set. For each $i \in I$, S_i has distribution $F(\theta_i)$ and $\theta_i \in \Theta$. It is assumed that $\boldsymbol{\theta} = \{\theta_i : i \in I\}$ is in a linear subspace Ω of the vector space V^I consisting of vectors $\mathbf{v} = \{v_i : i \in I\}$ such that $v_i \in V$ for $i \in I$.

Likelihood equations are readily derived for the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. This estimate is unique if it exists. Necessary and sufficient conditions are readily obtained for existence of $\hat{\boldsymbol{\theta}}$. These results are found in Section 2 by use of the general theory for exponential models developed by Berk (1972) and Barndorff-Nielsen (1973).

Asymptotic properties of $\hat{\boldsymbol{\theta}}$ are relatively easy to derive if the number of elements of I becomes large and the dimension of Ω is constant; however, results are much harder to obtain if both the number of elements of I and the dimension of Ω become large. Both situations are examined in Section 3 within the following context.

Let $\{I(n) : n \geqq 1\}$ be a sequence of distinct finite nonempty index sets with union J. For $n \geqq 1$, assume that $I(n) \subset I(n+1)$. For $i \in J$, let $\theta_i \in \Theta$ and let S_i be a random variable with distribution $F(\theta_i)$. Assume that the S_i, $i \in J$, are mutually independent.

For $n \geqq 1$, let Ω_n be a linear subspace of $V^{I(n)}$. Assume that the linear subspaces are compatible in the sense that for $m < n$,

$$\mathbf{x} = \{x_i : i \in I(m)\} \in \Omega_m$$

if and only if for some $\mathbf{y} = \{y_i : i \in I(n)\} \in \Omega_n$,

$$x_i = y_i \,, \qquad i \in I(m) \,.$$

For $\mathbf{x} = \{x_i : i \in J\} \in V^J$, let $\rho_n \mathbf{x} = \{x_i : i \in I(n)\}$. Let Ω_∞ consist of all $\mathbf{x} = \{x_i : i \in J\} \in V^J$ such that $\rho_n \mathbf{x} \in \Omega_n$ for all $n \geqq 1$.

For all $n \geqq 1$, let $\hat{\boldsymbol{\theta}}_n$ be the maximum likelihood estimate of $\boldsymbol{\theta}_n = \{\theta_i : i \in I(n)\}$ under the assumptions that $\boldsymbol{\theta}_n \in \Omega_n$ and that the S_i, $i \in I(n)$, have been observed. Let g be a linear functional on Ω_∞. If $\mathbf{x} \in \Omega_\infty$ and $\rho_n \mathbf{x} = \hat{\boldsymbol{\theta}}_n$, then $\hat{g}_n = g(\mathbf{x})$ is a maximum likelihood estimate of $g(\boldsymbol{\theta})$. If $g(\mathbf{x}) = 0$ whenever $\rho_n \mathbf{x} = \mathbf{0}$ and $\mathbf{x} \in \Omega_\infty$, then $\hat{g}_n$ is uniquely defined for $n \geqq m$.

In Section 3, conditions are provided under which the probability that $\hat{\boldsymbol{\theta}}_n$ exists approaches 1 as $n \to \infty$. Conditions are also provided under which $\hat{g}_n$ is asymptotically normal with asymptotic mean g. Other conditions are provided under which $\hat{g}_n$ is a consistent estimate of g. Expressions are derived for the asymptotic variance $\sigma_n(g)$ of $\hat{g}_n$, and conditions are provided under which the maximum likelihood estimate $\hat{\sigma}_n(g)$ of $\sigma_n(g)$ is a consistent estimate of $\sigma_n(g)$. Asymptotic confidence intervals are considered in cases in which $\hat{g}_n$ is asymptotically normal and asymptotically unbiased and $\hat{\sigma}_n(g)$ is a consistent estimate of $\sigma_n(g)$. Results of Section 3 rely on fixed-point theorems found in Kantorovich and Akilov (1964, pages 695–711).

The following examples help illustrate the problems considered in this paper.

EXAMPLE 1. *The Rasch model.* Rasch (1960, 1961) considers a family of models for use in educational tests in which subjects i, $1 \leqq i \leqq r$, have independent responses $S_{ij} \in T$ to test items j, $1 \leqq j \leqq c$. For some unknown $\alpha_i \in V$, $1 \leqq i \leqq r$, and $\beta_j \in V$, $i \leqq j \leqq c$, it is assumed that each S_{ij} has distribution $F(\alpha_i - \beta_j)$, $1 \leqq i \leqq r$, $1 \leqq j \leqq c$. Among cases considered by Rasch are the Poisson and the Bernoulli cases. In the Poisson case, $F(\theta)$, $\theta \in R$, is the Poisson distribution with mean e^θ. In the Bernoulli case, if S has distribution $F(\theta)$, $\theta \in R$, then

$$\Pr\{S = k\} = e^{\theta k}/(1 + e^\theta), \qquad k = 0 \text{ or } 1 .$$

In both examples, $V = R$ and $(x, y) = xy$ for real x and y, and $Y_{ij} = S_{ij}$. The parameter $\alpha_i - \alpha_k$ may be used to compare the relative abilities of subjects i and k, while $\beta_j - \beta_l$ may be used to compare relative difficulties of items j and l. To ensure that parameters have unique estimates, it is convenient to impose an arbitrary linear restriction. In this paper, it is assumed that $\beta_1 = 0$.

In cases in which the S_{ij} are Bernoulli random variables, considerable uncertainty exists concerning the appropriate manner for estimation of the parameters α_i and β_j. The conditional maximum likelihood approach of Andersen (1970, 1972) is impractical unless the number c of items is small. A similar problem is encountered by the Bock and Lieberman (1971) model in which the α_i are assumed to be independent identically distributed normal random variables. On the other hand, maximum likelihood estimates $\hat{\alpha}_i$ and $\hat{\beta}_j$ are readily computed. Unfortunately, the asymptotic properties of these estimates are little understood. Andersen (1973, pages 66–69) has shown that $\hat{\beta}_2$ is not a consistent estimate of β_2 in the case in which $c = 2$ and $r \to \infty$. Lord (1975) suggests that Andersen's result does not hold in the case $r \to \infty$ and $c \to \infty$. Lord's suggestion is supported by a recent Monte Carlo study of Wright and Douglas (1976). For the case $r = 500$ and $20 \leqq c \leqq 80$, they found $\hat{\beta}_j$ to be very similar to the conditional maximum likelihood estimate, and they found that the distribution of $\hat{\beta}_j$ was well approximated by use of standard asymptotic theory.

Some explanation for these results is provided in Section 3. Let $\{r_n : n \geqq 1\}$ be a strictly inceasing sequence, and let $\{c_n : n \geqq 1\}$ be an increasing sequence such that $c_n \to \infty$ as $n \to \infty$. Let $\{\alpha_i : i \geqq 1\}$ and $\{\beta_j : j \geqq 1\}$ be bounded sequences. Let $r_n \geqq c_n$ for each $n \geqq 1$. Let $\hat{\alpha}_{in}$ be the maximum likelihood estimate of α_i, $1 \leqq i \leqq r_n$, and let $\hat{\beta}_{jn}$ be the maximum likelihood estimate of β_j, $1 \leqq j \leqq c_n$, for the Rasch model for S_{ij}, $1 \leqq i \leqq r_n$, $1 \leqq j \leqq c_n$. If $c_n^{-1} \log r_n \to 0$, then for each i and j, $\hat{\alpha}_{in}$ converges in probability to α_i and $\hat{\beta}_{jn}$ converges in probability to β_j. Conditions are also given for asymptotic normality of maximum likelihood estimates.

EXAMPLE 2. *The Dempster model.* Dempster (1971) considers models in which for $1 \leqq i \leqq n$,

$$\theta_i = \sum_{j=1}^{p(n)} \beta_j x_{ij}$$

for some unknown $\beta_j \in V$, $1 \leqq j \leqq p(n)$, and some known x_{ij}, $1 \leqq i \leqq n$, $1 \leqq j \leqq p(n)$. Asymptotic properties of likelihood estimates $\hat{\beta}_{jn}$ of the β_j are not difficult to derive if $p(n)$ is constant for all n and $n \to \infty$. On the other hand, the situation is more difficult if $p(n) \to \infty$ and $x_{ij} = 0$ for $j > p(i)$. In this paper, asymptotic results are obtained for both cases. Under relatively general conditions, each $\hat{\beta}_{jn}$ is asymptotically normal and asymptotically unbiased.

2. Maximum likelihood equations. Derivation of the principal properties of maximum likelihood estimates is straightforward. Results follow from general theorems of Berk (1972) and Barndorff-Nielsen (1973), so details are omitted. To state results, let $Y_i = Y(S_i)$ for $i \in I$, and let $\mathbf{Y} = \{Y_i : i \in I\}$. Let

$$(\mathbf{y}, \mathbf{z})_I = \textstyle\sum_{i \in I} (y_i, z_i) , \qquad \mathbf{y}, \mathbf{z} \in V^I , \tag{2.1}$$

and let P be the orthogonal projection onto Ω with respect to $(\cdot, \cdot)_I$.

Let Θ_1 consist of all $\theta \in \Theta$ such that

$$\int \|Y(s)\| p(\theta, s)\, d\nu(s) < \infty ,$$

where $\|x\| = (x, x)^{\frac{1}{2}}$ for $x \in V$. As noted by Berk (1972), $\Theta^\circ \subset \Theta_1 \subset \Theta$. For $\theta \in \Theta_1$, let $E(\theta)$ be the expected value of $Y(S)$ when S is a random variable on T with distribution $F(\theta)$. Thus

$$E(\theta) = \int Y(s) p(\theta, s)\, d\nu(s) . \tag{2.2}$$

Given these definitions, the log-likelihood function $l(\mathbf{Y}, \theta)$ satisfies the equation

$$\begin{aligned} l(\mathbf{Y}, \boldsymbol{\theta}) &= \textstyle\sum_{i \in I} (Y_i, \theta_i) + \sum_{i \in I} \log a(\theta_i) \\ &= (\mathbf{Y}, \boldsymbol{\theta})_I + \textstyle\sum_{i \in I} \log a(\theta_i) \\ &= (P\mathbf{Y}, \boldsymbol{\theta})_I + \textstyle\sum_{i \in I} \log a(\theta_i) . \end{aligned} \tag{2.3}$$

In the language of Barndorff-Nielsen (1973) or Andersen (1974), $P\mathbf{Y}$ is a canonical statistic, $\boldsymbol{\theta}$ is a canonical parameter, and $P\mathbf{E}$ is a mean value parameter, where $\mathbf{E} = \{E(\theta_i) : i \in I\}$. The canonical parameter space $\Delta = \Omega \cap \Theta^I$ has nonempty interior, and the convex support $D = PC^I$ of the distribution of $P\mathbf{Y}$ has nonempty interior $D^\circ = P(C^\circ)^I$. The function $l(\mathbf{Y}, \cdot)$ is strictly concave on the convex set Δ.

Given these preliminary observations, the following theorems may be derived:

THEOREM 1. *Let $\hat{\boldsymbol{\theta}} = \{\hat{\theta}_i : i \in I\} \in \Omega$, and let $\hat{\theta}_i \in \Theta_1$ for $i \in I$. Let $\hat{\mathbf{E}} = \{E(\hat{\theta}_i) : i \in I\}$. Then $\hat{\theta}$ is a maximum likelihood estimate of θ if and only if*

$$P\hat{\mathbf{E}} = P\mathbf{Y} . \tag{2.4}$$

PROOF. See Berk (1972).

THEOREM 2. *Assume that $\Delta^\circ = (\Theta^\circ)^I \cap \Omega$ is equal to $\Theta_1{}^I \cap \Omega$. A maximum likelihood estimate $\hat{\theta}$ of Θ exists if and only if for some $z_i \in C^\circ$, $i \in I$, $\mathbf{z} = \{z_i : i \in I\}$ satisfies the equation $P\mathbf{z} = P\mathbf{Y}$.*

PROOF. Note that $\Delta^\circ = \Theta_1{}^I \cap \Omega$ if and only if the following condition is satisfied:

If $\mathbf{x}_t \in \Delta^\circ$, $t \geqq 0$, and $\mathbf{x}_t \to \mathbf{x} \notin \Delta^\circ$ as $t \to \infty$, then

$$\sum_{i \in I} ||E(x_{it})|| \to \infty .$$

Given this observation, the theorem follows from Barndorff-Nielsen (1973, Theorem 6.8). □

REMARK. The condition $\Delta^\circ = \Theta_1{}^I \cap \Omega$ holds if $\Theta^\circ = \Theta_1$, which is the case if Θ is open. Barndorff-Nielsen (1973) calls $\mathscr{F}$ a regular exponential family if Θ is open. The natural parameter space Θ is open if Y has a bounded range. This condition also holds for exponential families associated with the normal, gamma, Poisson, beta and noncentral t distributions.

THEOREM 3. *Assume that $\Delta^\circ = \Theta_1{}^I \cap \Omega$. A maximum likelihood estimate $\hat{\theta}$ of θ exists if and only if no $x \in \Omega$ satisfies the following conditions*:

$$x_i = 0 \quad \text{for} \quad i \in I \quad \text{such that} \quad Y_i \in C^\circ , \tag{2.5}$$

$$(x_i, Y_i - c) \geqq 0 , \quad c \in C^\circ , \quad \text{for} \quad i \in I \quad \text{such that} \quad Y_i \notin C^\circ , \tag{2.6}$$

$$(x_j, Y_j - c) > 0 , \quad c \in C^\circ , \quad \text{for some} \quad j \in I \quad \text{such that} \quad Y_j \notin C^\circ . \tag{2.7}$$

PROOF. Assume $\mathbf{x} \in \Omega$ satisfies (2.5), (2.6) and (2.7). If $z_i \in C^\circ$ for $i \in I$, then $\mathbf{z} = \{z_i : i \in I\}$ satisfies the inequality

$$(\mathbf{x}, \mathbf{Y} - \mathbf{z})_I = \sum_{i \in I} (x_i, Y_i - z_i) > 0 .$$

Therefore, $P\mathbf{z} \neq P\mathbf{Y}$. By Theorem 2, $\hat{\boldsymbol{\theta}}$ does not exist.

Assume now that $\hat{\boldsymbol{\theta}}$ does not exist. Let $\Omega^\perp$, the orthogonal complement of Ω, be defined as $\{\mathbf{x} \in V^I : (\mathbf{x}, \mathbf{y})_I = 0, \mathbf{y} \in \Omega\}$. Let $A = \{\mathbf{z} - \mathbf{Y} : \mathbf{z} \in (C^\circ)^I\}$. By Theorem 2, A and $\Omega^\perp$ are disjoint. Since A is open and convex, it follows from Rockafeller (1970, page 96) that for some $\mathbf{x} \in \Omega$, $(\mathbf{x}, \mathbf{y})_I < 0$ for all $\mathbf{y} \in A$.

Consider an $i \in I$. Let $z_{kt} \in C^\circ$, $k \in I$, $t \geqq 0$, be defined so that as $t \to \infty$,

$$z_{kt} \to Y_k , \qquad k \neq i .$$

Let

$$z_{it} = c \in C^\circ , \qquad t \geqq 0 .$$

Such a sequence of $\mathbf{z}_t = \{z_{kt} : k \in I\}$ can be found since Y has range in C. Since

$$\sum_{k \in I} (x_k, z_{kt} - Y_k) < 0 , \qquad t \geqq 0 ,$$

it follows that $(x_i, c - Y_i) \leqq 0$. Thus

$$(x_i, Y_i - c) \geqq 0 , \qquad c \in C^\circ .$$

If $Y_i \notin C^\circ$, then (2.6) holds. If $Y_i \in C^\circ$, then for some $\varepsilon > 0$, $Y_i + \varepsilon x_i \in C^\circ$ and $-\varepsilon(x_i, x_i) \geqq 0$. Thus (2.5) holds. Since

$$\sum_{i \in I} (x_i, c_i - Y_i) < 0 , \qquad c_i \in C^\circ \quad \text{for} \quad i \in I ,$$

(2.7) must hold for some $j \in I$ such that $Y_j \notin C^\circ$. □

THEOREM 4. *If there exists a maximum likelihood estimate $\hat{\theta}$ of θ, then $\hat{\theta}$ is uniquely defined.*

PROOF. The result follows since $l(\mathbf{Y}, \cdot)$ is strictly concave. See Berk (1972). □

To illustrate results, consider the following examples.

EXAMPLE 1. *The Rasch model* (*continued*). Let $\hat{E}_{ij} = E(\hat{\theta}_{ij})$, $\hat{\theta}_{ij} = \hat{\alpha}_i - \hat{\beta}_j$,

$$Y_{i+} = \textstyle\sum_{j=1}^{c} Y_{ij}\,, \qquad 1 \leqq i \leqq r\,,$$
$$Y_{+j} = \textstyle\sum_{i=1}^{r} Y_{ij}\,, \qquad 1 \leqq i \leqq c\,,$$
$$\hat{E}_{i+} = \textstyle\sum_{j=1}^{c} \hat{E}_{ij}\,, \qquad 1 \leqq i \leqq r\,,$$
$$\hat{E}_{+j} = \textstyle\sum_{i=1}^{r} \hat{E}_{ij}\,, \qquad 1 \leqq j \leqq c\,.$$

The subspace Ω in this example consists of all $\mathbf{x} = \{x_{ij}: 1 \leqq i \leqq r, 1 \leqq j \leqq c\}$ such that

$$x_{ij} = y_i + z_j\,, \qquad 1 \leqq i \leqq r,\ 1 \leqq j \leqq c\,,$$

for some $y_i \in V$, $1 \leqq i \leqq r$, and $z_j \in V$, $1 \leqq j \leqq c$. By (2.4), if $y_i \in V$, $1 \leqq i \leqq r$, and $z_j \in V$, $1 \leqq j \leqq c$, then

$$\begin{aligned}\textstyle\sum_{i=1}^{r} \sum_{j=1}^{c} (y_i + z_j, \hat{E}_{ij}) &= \textstyle\sum_{i=1}^{r} (y_i, \hat{E}_{i+}) + \sum_{j=1}^{c} (z_j, \hat{E}_{+j}) \\ &= \textstyle\sum_{i=1}^{r} (y_i, Y_{i+}) + \sum_{j=1}^{c} (z_j, Y_{+j})\,.\end{aligned}$$

Equivalently,

$$\hat{E}_{i+} = Y_{i+}\,, \qquad 1 \leqq i \leqq r\,,$$
$$\hat{E}_{+j} = Y_{+j}\,, \qquad 1 \leqq j \leqq c\,.$$

To obtain conditions for existence of $\hat{\boldsymbol{\theta}}$, note that $\Theta_1^I \cap \Omega$ includes all $\mathbf{x} = \{x_{ij}: 1 \leqq i \leqq r, 1 \leqq j \leqq c\}$ such that $x_{ij} = \theta$, $1 \leqq i \leqq r$, $1 \leqq j \leqq c$, and $\theta \in \Theta_1$. Therefore, $\Theta_1^I \cap \Omega = (\Theta^\circ)^I \cap \Omega$ if and only if $\Theta_1 = \Theta^\circ$. By Theorem 2, if $\Theta_1 = \Theta^\circ$, then $\hat{\boldsymbol{\theta}}$ exists if and only if for some $z_{ij} \in C^\circ$, $1 \leqq i \leqq r$, $1 \leqq j \leqq c$,

$$z_{i+} = Y_{i+}\,, \qquad 1 \leqq i \leqq r\,,$$
$$z_{+j} = Y_{+j}\,, \qquad 1 \leqq j \leqq c\,.$$

Thus a necessary condition for $\hat{\boldsymbol{\theta}}$ to exist is that

$$Y_{i\bullet} = \frac{1}{c} Y_{i+} \in C^\circ\,, \qquad 1 \leqq i \leqq r\,,$$

$$Y_{\bullet j} = \frac{1}{r} Y_{+j} \in C^\circ\,, \qquad 1 \leqq j \leqq c\,.$$

In the Poisson case, this condition on $Y_{i\bullet}$ and $Y_{\bullet j}$ is also sufficient. Note that $C = [0, \infty)$ and $C^\circ = (0, \infty)$. If

$$Y_{i\bullet} > 0\,, \qquad 1 \leqq i \leqq r\,, \tag{2.8}$$

$$Y_{\bullet j} > 0\,, \qquad 1 \leqq j \leqq c\,, \tag{2.9}$$

then it is well known that

$$\hat{E}_{ij} = Y_{i+} Y_{+j} / Y_{++}\,,$$

where

$$Y_{++} = \sum_{i=1}^{r} Y_{i+} .$$

As already noted, $\hat{\boldsymbol{\theta}}$ cannot exist unless (2.8) and (2.9) hold.

In the Bernoulli case, the necessary and sufficient condition for existence of $\hat{\boldsymbol{\theta}}$ is more complex than in the Poisson case. Here $C = [0, 1]$ and $C^\circ = (0, 1)$. The estimate $\hat{\boldsymbol{\theta}}$ fails to exist if and only if sets A, B, C and D exist such that the following conditions hold:

$$A \neq \emptyset \quad \text{and} \quad C \neq \emptyset , \qquad \text{or} \qquad B \neq \emptyset \quad \text{and} \quad D \neq \emptyset , \tag{2.8}$$

$$A \cap B = C \cap D = \emptyset , \tag{2.9}$$

$$A \cup B = \bar{r} , \tag{2.10}$$

$$C \cup D = \bar{c} , \tag{2.11}$$

$$\begin{aligned} Y_{ij} &= 0 , \qquad i \in A , \quad j \in C , \\ &= 1 , \qquad i \in B , \quad j \in D . \end{aligned} \tag{2.12}$$

In these conditions, if $s \geqq 1$ is an integer, then $\bar{s}$ consists of all integers i, $1 \leqq i \leqq s$.

To verify this condition, assume first that A, B, C and D satisfy (2.8)—(2.12). Let

$$x_{ij} = g_i + h_j , \qquad 1 \leqq i \leqq r,\ 1 \leqq j \leqq c ,$$

where

$$\begin{aligned} g_i &= -1 , && i \in A , \\ &= 1 , && i \in B , \\ h_j &= -1 , && j \in C , \\ &= 1 , && j \in D . \end{aligned}$$

Note that for all $d \in (0, 1)$, $i \in \bar{r}$, and $j \in \bar{c}$,

$$x_{ij}(Y_{ij} - d) \geqq 0 ,$$

with strict inequality for $i \in A$ and $j \in C$ or for $i \in B$ and $j \in D$. Since no Y_{ij} can be in C°, Theorem 3 implies that $\hat{\boldsymbol{\theta}}$ does not exist.

Assume now that $\hat{\boldsymbol{\theta}}$ does not exist. By Theorem 3, there exists g_i, $1 \leqq i \leqq r$, and h_j, $1 \leqq j \leqq c$, such that

$$(g_i + h_j)(Y_{ij} - d) \geqq 0 , \qquad d \in (0, 1),\ i \in \bar{r} \ \text{ and } \ j \in \bar{c} ,$$

with strict inequality for some i and j. Thus $g_i + h_j \geqq 0$ if $Y_{ij} = 1$ and $g_i + h_j \leqq 0$ if $Y_{ij} = 0$. For some k, $1 \leqq k \leqq r$, and some l, $1 \leqq l \leqq c$, $g_k + h_l \neq 0$. Without loss of generality, let k and l be chosen so that

$$\begin{aligned} g_k + h_l &> 0 , && \\ g_i + h_l &\leqq 0 , && g_i < g_k , \\ g_k + h_j &\leqq 0 , && h_j < h_l . \end{aligned}$$

Let $A = \{i \in \bar{r} : g_i < g_k\}$, $B = \{i \in \bar{r} : g_i \geqq g_k\}$, $C = \{j \in \bar{c} : h_j < h_l\}$, and $D = \{j \in \bar{c} : h_j \geqq h_l\}$. Clearly (2.8)—(2.11) hold. To verify (2.12), let $i \in A$ and $j \in C$. Note that

$$g_i + h_j = (g_i + h_l) + (h_j - h_l) < 0 ,$$

so that $Y_{ij} = 0$. On the other hand, if $i \in B$ and $j \in D$, then $g_i + h_j \geqq g_k + h_l > 0$. Thus $Y_{ij} = 1$.

This necessary and sufficient condition for the Bernoulli case implies that $\hat{\boldsymbol{\theta}}$ does not exist if $Y_{i+} = 0$ or $Y_{i+} = c$ for some i, $1 \leqq i \leqq r$, or if $Y_{+j} = 0$ or $Y_{+j} = r$ for some j, $1 \leqq j \leqq c$. These cases essentially involve inability to estimate subject abilities or item difficulties.

EXAMPLE 2. *The Dempster model* (*continued*). The maximum likelihood equations for this model may be written

$$\sum_{i=1}^{n} x_{ij} \hat{E}_i = \sum_{i=1}^{n} x_{ij} Y_i , \qquad 1 \leqq j \leqq p(n) ,$$

as noted in Dempster (1971). If $\Theta_1 = \Theta^\circ$, then $\hat{\boldsymbol{\theta}}$ exists if and only if for some $z_i \in C^\circ$, $1 \leqq i \leqq n$,

$$\sum_{i=1}^{n} x_{ij} z_i = \sum_{i=1}^{n} x_{ij} Y_i , \qquad 1 \leqq j \leqq p(n) .$$

3. Asymptotic properties. In this section, asymptotic behavior of maximum likelihood estimates is examined. As noted in Section 1, a sequence $\{I(n) : n \geqq 1\}$ of distinct finite nonempty index sets is given such that $I(n) \subset I(n+1)$ for $n \geqq 1$ and such that

$$\bigcup_{n=1}^{\infty} I(n) = J .$$

The following conditions are assumed to hold: For $i \in J$, $\theta_i \in \Theta$ and S_i has distribution $F(\theta_i)$. The S_i, $i \in J$, are mutually independent. For $n \geqq 1$, $\boldsymbol{\theta}_n = \{\theta_i : i \in I(n)\} \in \Omega_n$, a linear subspace of $V^{I(n)}$. If $m < n$ and $\mathbf{x} = \{x_i : i \in I(m)\} \in \Omega_m$, then for some $\mathbf{y} = \{y_i : i \in I(n)\} \in \Omega_n$,

$$x_i = y_i , \qquad i \in I(m) .$$

The space Ω_∞ consists of $\mathbf{x} \in V^J$ such that $\rho_n \mathbf{x} = \{x_i : i \in I(n)\} \in \Omega_n$ for all $n \geqq 1$.

For $n \geqq 1$, let $\hat{\boldsymbol{\theta}}_n = \{\hat{\theta}_{in} : i \in I(n)\}$ be the maximum likelihood estimate of $\boldsymbol{\theta}_n$ under the assumptions that $\boldsymbol{\theta}_n \in \Omega_n$ and that the S_i, $i \in I(n)$, have been observed. Let $\Omega_\infty{}^*$ be the space of linear functionals of Ω_∞, and let $\Omega_{n\infty}^*$ consist of all $g \in \Omega_\infty{}^*$ such that

$$g(\mathbf{x}) = 0 \quad \text{if} \quad \mathbf{x} \in \Omega_\infty \quad \text{and} \quad \rho_n \mathbf{x} = \mathbf{0} .$$

Note that if $\hat{\boldsymbol{\theta}}_n$ exists, $g \in \Omega_{n\infty}^*$, and $\rho_n \mathbf{x} = \hat{\boldsymbol{\theta}}_n$, then $\hat{g}_n = g(\mathbf{x})$ is the unique maximum likelihood estimate of $g(\boldsymbol{\theta})$. If $m < n$, then $\Omega_{m\infty}^* \subset \Omega_{n\infty}^*$.

For $g \in \Omega_\infty{}^*$, let $n(g)$ be the smallest $n \geqq 1$ such that $g \in \Omega_{n\infty}^*$. If no such n exists, let $n(g) = \infty$. The major objective in this section is determination of conditions under which

$$\frac{\hat{g}_n - g(\boldsymbol{\theta})}{\sigma_n(g)} \to_{\mathscr{D}} N(0, 1) . \tag{3.1}$$

Here $g \in \Omega_\infty^*$, $n(g) < \infty$, $\sigma_n(g)$ is the asymptotic standard deviation of $\hat{g}_n$, and $\rightarrow_{\mathscr{D}}$ signifies convergence in distribution. Conditions are also given under which

$$\sigma_n(g) \rightarrow 0 , \tag{3.2}$$

and the maximum likelihood estimate $\hat{\sigma}_n(g)$ of $\sigma_n(g)$ satisfies

$$\hat{\sigma}_n(g)/\sigma_n(g) \rightarrow_P 1 , \tag{3.3}$$

where $\rightarrow_P$ signifies convergence in probability. These results permit construction of two-sided level-$(1-\alpha)$ asymptotic confidence intervals

$$U_\alpha(g) = (\hat{g}_n - Z_{\alpha/2}\hat{\sigma}_n(g), \hat{g}_n + Z_{\alpha/2}\hat{\sigma}_n(g)) , \tag{3.4}$$

where $Z_{\alpha/2}$ is the upper-$(\alpha/2)$ point of the standard normal distribution. Note that (3.1) and (3.3) imply that

$$\Pr\{g(\boldsymbol{\theta}) \in U_\alpha(g)\} \rightarrow 1 - \alpha , \qquad 0 < \alpha < 1 .$$

Also note that (3.2) implies that $\hat{g}_n$ is a consistent estimate of $g(\boldsymbol{\theta})$.

To define $\sigma_n(g)$ and $\hat{\sigma}_n(g)$, the covariance operator $\Sigma(\theta)$ must be defined for $\boldsymbol{\theta} \in \Theta_2$. Here

$$\Theta_2 = \{\theta \in \Theta : \int ||Y(s)||^2 \exp(\theta, Y(s))\, d\nu(s) < \infty\}$$

satisfies the inclusion relationships $\Theta^\circ \subset \Theta_2 \subset \Theta_1 \subset \Theta$. For $\theta \in \Theta_2$, let $\Sigma(\theta)$ be the covariance operator of a random vector $Y(S)$, where $S \in T$ has distribution $F(\theta)$. Thus

$$(v, \Sigma(\theta)w) = \int (v, Y(s) - E(\theta))(w, Y(s) - E(\theta))p(\theta, s)\, d\nu(s) , \quad v, w \in V . \tag{3.6}$$

As noted in Berk (1972), $\Sigma(\theta)$ is positive definite.

For $i \in J$, let $\Sigma_i = \Sigma(\theta_i)$. For $\mathbf{y}, \mathbf{z} \in V^{I(n)}$, let

$$[\mathbf{y}, \mathbf{z}]_n = \textstyle\sum_{i \in I(n)} (y_i, \Sigma_i z_i) \tag{3.7}$$

and

$$||\mathbf{y}||_n^2 = [\mathbf{y}, \mathbf{y}]_n . \tag{3.8}$$

If $g \in \Omega_{n\infty}^*$, then the asymptotic standard deviation $\sigma_n(g)$ of $\hat{g}_n$ is the supremum of $|g(\mathbf{x})|$ for $\mathbf{x} \in \Omega_\infty$ such that $||\rho_n \mathbf{x}||_n \leqq 1$. Note that $\sigma_n(g) < \infty$ if $g \in \Omega_{n\infty}^*$.

If $\hat{\boldsymbol{\theta}}_n$ exists and $\hat{\theta}_{in} \in \Theta_2$ for $i \in I(n)$, then $\Sigma_{in} = \Sigma(\hat{\theta}_{in})$ is uniquely defined for $i \in I(n)$. Let

$$|\mathbf{y}|_n^2 = \textstyle\sum_{i \in I(n)} (y_i, \Sigma_{in} y_i) , \qquad \mathbf{y} \in V^{I(n)} . \tag{3.9}$$

The maximum likelihood estimate $\hat{\sigma}_n(g)$ of $\sigma_n(g)$ is the supremum of $|g(\mathbf{x})|$ for $\mathbf{x} \in \Omega_\infty$ such that $|\rho_n \mathbf{x}|_n \leqq 1$. Computation of $\sigma_n(g)$ and $\hat{\sigma}_n(g)$ is discussed later in this section.

To provide conditions for (3.1), (3.2) and (3.3) to hold, consider a subset A of Ω_∞^*. Let $A_n = A \cap \Omega_{n\infty}^*$ for $n \geqq 1$. Assume that if $\mathbf{x} \in \Omega_\infty$ and $g(\mathbf{x}) = 0$ for all $g \in A_n$, then $\rho_n \mathbf{x} = \mathbf{0}$. Let $||\cdot||_n(A)$ be the norm on Ω_n defined so that for

$\mathbf{y} \in \Omega_n$, $||\mathbf{y}||_n(A)$ is the smallest nonnegative number such that

$$(3.10)\qquad |g(\mathbf{x})| \leqq ||\mathbf{y}||_n(A)\sigma_n(g)\,, \qquad \mathbf{x} \in \Omega\,, \quad g \in A_n\,, \quad \rho_n \mathbf{x} = \mathbf{y}\,.$$

Note that $||\mathbf{y}||_n(A) \leqq ||\mathbf{y}||_n(\Omega_\infty^*) = ||\mathbf{y}||_n$.

Let $\mathbf{Y}_n = \{Y_i : i \in I(n)\}$, and let

$$(3.11)\qquad \mathbf{E}_n = \{E_i : i \in I(n)\} = \{E(\theta_i) : i \in I(n)\}$$

denote the expected value of $\mathbf{Y}_n$. Let $\mathbf{Z}_n \in \Omega_n$ be defined by the equation

$$(3.12)\qquad [\mathbf{x}, \mathbf{Z}_n] = \textstyle\sum_{i \in I(n)} (x_i, Y_i - E_i)\,, \qquad \mathbf{x} \in \Omega_n\,.$$

Much of the analysis of this section depends on the possibility that $\hat{\boldsymbol{\theta}}_n$ may be approximated by $\boldsymbol{\theta}_n + \mathbf{Z}_n$. The accuracy of this approximation depends on $\mathbf{U}_n(\cdot, \cdot)$. Here for $\mathbf{y} \in \Theta_2^{I(n)} \cap \Omega_n$ and $\mathbf{z} \in \Omega_n$, $\mathbf{U}_n(\mathbf{y}, \mathbf{z}) \in \Omega_n$ is defined by the equation

$$(3.13)\qquad [\mathbf{x}, \mathbf{U}_n(\mathbf{y}, \mathbf{z})]_n = \textstyle\sum_{i \in I(n)} (x_i, [\Sigma(y_i) - \Sigma_i] z_i)\,.$$

The fundamental condition required in this section is the following:

CONDITION 1. *A subset $A \subset \Omega_\infty^*$ and constants $n' \geqq 1$, $e > 0$, $d_n \geqq 0$, $n \geqq n'$, and $f_n \geqq e$, $n \geqq n'$, exist such that the following conditions hold for $n \geqq n'$:*

(a) *If $\mathbf{x} \in \Omega_\infty$ and $g(\mathbf{x}) = 0$ for $g \in A_n = A \cap \Omega_{n\infty}^*$, then $\rho_n \mathbf{x} = \mathbf{0}$.*

(b) *If $\mathbf{y}, \mathbf{z} \in \Omega_n$ and $||\mathbf{y} - \boldsymbol{\theta}_n||_n(A) \leqq f_n$, then*

$$(3.14)\qquad y_i \in \Theta^\circ\,, \qquad i \in I(n)\,,$$

and

$$(3.15)\qquad ||\mathbf{U}_n(\mathbf{y}, \mathbf{z})||_n(A) \leqq d_n ||\mathbf{y} - \boldsymbol{\theta}_n||_n(A) ||\mathbf{z}||_n(A)\,.$$

(c) *As $n \to \infty$, $\Pr\{||\mathbf{Z}_n||_n(A) < \frac{1}{2} f_n\} \to 1$.*

(d) *As $n \to \infty$, $d_n f_n \to 0$.*

Procedures for verification of Condition 1 are discussed later in this section. Given Condition 1, the following theorems may be derived.

THEOREM 5. *Assume Condition 1 holds. As $n \to \infty$, the probability approaches 1 that $\hat{\boldsymbol{\theta}}_n$ exists.*

PROOF. Fixed-point theorems of Kantorovich and Akilov (1964, pages 695–711) are used to establish this result. The initial step required is construction of a function M_n on $\Delta_{n1} = \Omega_n \cap \Theta_1^{I(n)}$ with a fixed point $\hat{\boldsymbol{\theta}}_n$.

Let $L_n(\mathbf{y}) \in \Omega_n$ be defined for $\mathbf{y} \in \Delta_{n1}$ by the equation

$$[\mathbf{z}, L_n(\mathbf{y})]_n = \textstyle\sum_{i \in I(n)} (z_i, Y_i - E(y_i))\,, \qquad \mathbf{z} \in \Omega_n\,,$$

and let

$$M_n(\mathbf{y}) = \mathbf{y} + L_n(\mathbf{y})\,.$$

Note that M_n has a fixed point $\mathbf{w} \in \Delta_{n1}$ if and only if

$$\textstyle\sum_{i \in I(n)} (z_i, Y_i - E(w_i)) = 0\,, \qquad \mathbf{z} \in \Omega_n\,.$$

By Theorem 1 and Theorem 4, $\mathbf{w}$ exists if and only if $\hat{\boldsymbol{\theta}}_n$ exists. If $\hat{\boldsymbol{\theta}}_n$ exists, $\mathbf{w} = \hat{\boldsymbol{\theta}}_n$.

The fixed-point theorem used here requires construction of the sequence $\{\mathbf{v}_{nk} : k \geqq 0\}$. Here

$$\mathbf{v}_{n0} = \boldsymbol{\theta}_n ,$$
$$\mathbf{v}_{n(k+1)} = M_n(\mathbf{v}_{nk}) , \qquad k \geqq 0 .$$

Note that

$$\mathbf{v}_{n1} = \boldsymbol{\theta}_n + \mathbf{Z}_n .$$

To determine whether the sequence is well defined and converges to $\hat{\boldsymbol{\theta}}_n$, the differential $dM_{n\mathbf{y}}$ of M_n at $\mathbf{y} \in \Delta_n{}^\circ = (\Theta^\circ)^{I(n)} \cap \Omega_n$ must be examined. This differential satisfies the equation

$$dM_{n\mathbf{y}}(\mathbf{z}) = -\mathbf{U}_n(\mathbf{y}, \mathbf{z}) , \qquad \mathbf{z} \in \Omega_n .$$

To verify this claim, note, as in Berk (1972), that $E(\cdot)$ has differential $\mathbf{\Sigma}(\theta)$ at $\theta \in \Theta^\circ$. Thus for $\mathbf{y} + \mathbf{z} \in \Delta_n{}^\circ$,

$$\begin{aligned}[\mathbf{x}, M_n(\mathbf{y} + \mathbf{z}) - M_n(\mathbf{y})]_n &= \textstyle\sum_{i \in I(n)} [(x_i, \mathbf{\Sigma}_i z_i) + (x_i, E(y_i) - E(y_i + z_i))] \\ &= -[\mathbf{x}, \mathbf{U}_n(\mathbf{y}, \mathbf{z})]_n + o(\mathbf{z}) ,\end{aligned}$$

where $o(\mathbf{z})/||\mathbf{z}||_n \to 0$ as $||\mathbf{z}||_n \to 0$.

Let $||dM_n\mathbf{y}||_n(A)$ denote the smallest nonnegative number such that

$$||dM_{n\mathbf{y}}(z)||_n(A) \leqq ||dM_{n\mathbf{y}}||_n(A)||\mathbf{z}||_n(A) , \qquad \mathbf{z} \in \Omega_n .$$

Then

$$||dM_{n\mathbf{y}}||_n(A) \leqq d_n||\mathbf{y} - \boldsymbol{\theta}_n||_n(A) , \quad \mathbf{y} \in \Omega_n , \qquad ||\mathbf{y} - \boldsymbol{\theta}_n||_n(A) \leqq f_n .$$

As shown in Kantorovich and Akilov (1964, pages 695–711), if $||\mathbf{Z}_n||_n(A) < \frac{1}{2}f_n$ and $d_n||\mathbf{Z}_n||_n(A) < \frac{1}{2}$, then $\hat{\boldsymbol{\theta}}_n$ exists. Since $\Pr\{||\mathbf{Z}_n||_n(A) < \frac{1}{2}f_n\} \to 1$, the probability approaches 1 that $\hat{\boldsymbol{\theta}}_n$ exists.

THEOREM 6. *Assume that Condition* 1 *holds. As $n \to \infty$, the probability approaches* 1 *that*

$$|||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n||_n(A) - ||\mathbf{Z}_n||_n(A)|^2 \leqq d_n[||\mathbf{Z}_n||_n(A)]^2$$

and

$$||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n - \mathbf{Z}_n||_n(A) \leqq d_n[||\mathbf{Z}_n||_n(A)]^2 .$$

PROOF. Define $\mathbf{v}_{nk}$, $k \geqq 0$, as in the proof of Theorem 5. Let $z_n = ||\mathbf{Z}_n||_n(A)$. Let $t_{n0} = 0$, and for $k \geqq 0$, let

$$t_{n(k+1)} = z_n + \tfrac{1}{2}d_n t_{nk}^2 .$$

Let

$$r_n = 2z_n/[1 + (1 - 2z_n d_n)^{\frac{1}{2}}] .$$

As Kantorovich and Akilov (1964, pages 695–711) have shown, if $z_n < \frac{1}{2}f_n$ and $z_n d_n < \frac{1}{2}$, then

$$||\mathbf{v}_{nk} - \hat{\boldsymbol{\theta}}_n||_n(A) \leqq r_n - t_{nk} , \qquad k \geqq 0 . \tag{3.16}$$

For $k = 0$, (3.16) implies that

$$||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n||_n(A) \leqq r_n .$$

For $k = 1$, (3.16) implies that

$$||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n - \mathbf{Z}_n||_n(A) \leqq r_n - z_n .$$

Thus (3.16) implies

$$|||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n||_n(A) - ||\mathbf{Z}_n||_n(A)| \leqq r_n - z_n .$$

Note that

$$r_n - z_n = \frac{z_n[1 - (1 - 2z_n d_n)^{\frac{1}{2}}]}{[1 + (1 - 2z_n d_n)^{\frac{1}{2}}]} .$$

Examination of derivatives of the function $[1 - (1 - x)^{\frac{1}{2}}]/[1 + (1 - x)^{\frac{1}{2}}]$ for $0 \leqq x < \frac{1}{2}$ shows that

$$r_n - z_n \leqq d_n z_n^2$$

whenever $d_n z_n < \frac{1}{2}$. Since $\Pr\{z_n < \frac{1}{2}f_n\} \to 1$ and $d_n f_n \to 0$, the conclusions of the theorem follow. □

3.1. *Consistency of estimates.* To obtain consistency results, a further condition is required to supplement Condition 1. This new condition is defined in terms of b_n, the smallest nonnegative number such that

$$||x_i|| \leqq b_n||\mathbf{x}||_n , \qquad i \in I(n),\ \mathbf{x} \in \Omega_n .$$

The coefficient b_n is positive if and only if Ω_n contains a nonzero element. Thus b_n is the maximum asymptotic standard deviation $\sigma_n(g)$ for $g \in \Omega_\infty^*$ such that for some $i \in I(n)$ and $c \in V$, $||c|| = 1$ and

$$g(\mathbf{x}) = (c, x_i) , \qquad \mathbf{x} \in \Omega_\infty .$$

Consistency results for $\hat{g}_n$ require Condition 1 and the condition $d_n b_n \to 0$. The following theorems may be derived.

THEOREM 7. *Let $g \in \Omega_\infty^*$ and $n(g) < \infty$. Then for some $\gamma(g) \geqq 0$, $\sigma_n(g) \leqq \gamma(g)b_n$ for $n \geqq n(g)$. If $b_n \to 0$ as $n \to \infty$, then $\sigma_n(g) \to 0$.*

PROOF. For some unique $\mathbf{c} \in \Omega_{n(g)}$,

$$g(\mathbf{x}) = \textstyle\sum_{i \in I(n(g))} (c_i, x_i) , \qquad \mathbf{x} \in \Omega .$$

Let

$$\gamma(g) = \textstyle\sum_{i \in I(n(g))} ||c_i|| .$$

Then

$$|g(\mathbf{x})| \leqq b_n \gamma(g)||\mathbf{x}||_n , \qquad \mathbf{x} \in \Omega_\infty,\ n \geqq n(g) .$$

Thus $\sigma_n(g) \leqq \gamma(g)b_n$, and $\sigma_n(g) \to 0$ if $b_n \to 0$. □

THEOREM 8. *Assume Condition 1 holds and $b_n f_n \to 0$ as $n \to \infty$. Let $g \in \Omega_\infty^*$ and $n(g) < \infty$. Then $\hat{g}_n \to_P g(\boldsymbol{\theta})$ as $n \to \infty$.*

PROOF. For $n \geqq n(g)$, let $\mathbf{x}^{(n)} \in \Omega_\infty$ satisfy the equation $\rho_n \mathbf{x}^{(n)} = \hat{\boldsymbol{\theta}}_n$. Then $\hat{g}_n = g(\mathbf{x}^{(n)})$ and $\hat{g}_n - g(\boldsymbol{\theta}) = g(\mathbf{x}^{(n)} - \boldsymbol{\theta})$.

Let $\tau_n(g, A)$ be the supremum of $|g(\mathbf{y})|$ for $\mathbf{y} \in \Omega_\infty$ such that $||\rho_n \mathbf{y}||_n(A) \leqq 1$. By (3.10)

$$|\hat{g}_n - g(\boldsymbol{\theta})| \leqq \tau_n(g, A)||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n||_n(A) .$$

Given the definition of $||\cdot||_n(A)$ and Theorem 7, for some $\delta(g, A) \geqq 0$,

$$\tau_n(g, A) \leqq \delta(g, A)b_n , \quad n \geqq n(g) .$$

By Theorem 6 and the condition $f_n b_n \to 0$, it follows that $|\hat{g}_n - g(\boldsymbol{\theta})| \to_P 0$. □

THEOREM 9. *Assume Condition* 1 *holds, and assume that for some* $d_n' \geqq 0, n \geqq 1$,

$$|\textstyle\sum_{i \in I(n)} (x_i, [\Sigma(y_i) - \Sigma_i]x_i)| \leqq d_n'||\mathbf{x}||_n^2||\mathbf{y} - \boldsymbol{\theta}_n||_n(A) , \qquad \mathbf{x}, \mathbf{y} \in \Omega_n ,$$
$$||\mathbf{y} - \boldsymbol{\theta}_n||_n(A) \leqq f_n .$$

Assume that $d_n' f_n \to 0$. *Let* $g \in \Omega_\infty^*$ *and* $n(g) < \infty$. *Assume* $g(x) \neq 0$ *for some* $\mathbf{x} \in \Omega_\infty$. *Then*

$$\hat{\sigma}_n(g)/\sigma_n(g) \to_P 1 .$$

REMARK. If $A = \Omega_\infty^*$, then one may let $d_n' = d_n$, so that Condition 1 implies that $d_n' f_n \to 0$.

PROOF. Let $W_n = ||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n||_n(A)$. If $W_n \leqq f_n$ and $\mathbf{x} \in \Omega_\infty$, then

$$|\textstyle\sum_{i \in I(n)} (x_i, \Sigma_{in} x_i) - \sum_{i \in I(n)} (x_i, \Sigma_i x_i)| \leqq d_n' W_n ||\mathbf{x}||_n^2 .$$

Since $g(\mathbf{x}) \neq 0$ for some $\mathbf{x} \in \Omega$, $0 < \sigma_n(g) < \infty$ for $n \geqq n(g)$. The definitions of $\sigma_n(g)$ and $\hat{\sigma}_n(g)$ imply that

$$|\sigma_n^2(g)/\hat{\sigma}_n^2(g) - 1| \leqq d_n' W_n ,$$

provided $W_n \leqq f_n$. By Theorem 6, $\hat{\sigma}_n(g)/\sigma_n(g) \to_P 1$. □

3.2. *Asymptotic normality.* Asymptotic normality results are most readily derived if $g \in A$, $n(g) < \infty$, and g is not identically zero. In other cases, results are harder to obtain. The following theorems are available.

THEOREM 10. *Assume Condition* 1 *holds and* $d_n f_n^2 \to 0$. *Let* $g \in A$ *and* $n(g) < \infty$. *Assume* $g(\mathbf{x}) \neq 0$ *for some* $\mathbf{x} \in \Omega_\infty$. *Then*

$$[\hat{g}_n - g(\boldsymbol{\theta})]/\sigma_n(g) \to_{\mathscr{D}} N(0, 1) .$$

PROOF. For $n \geqq n(g)$, let $\mathbf{c}_n \in \Omega_n$ be defined by the equation

$$g(\mathbf{x}) = [\mathbf{c}_n, \rho_n \mathbf{x}]_n , \qquad \mathbf{x} \in \Omega_\infty .$$

Thus

$$\hat{g}_n - g(\boldsymbol{\theta}) = [\mathbf{c}_n, \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n]_n = [\mathbf{c}_n, \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n - \mathbf{Z}_n]_n + [\mathbf{c}_n, \mathbf{Z}_n] .$$

Since $\sigma_n(g)$ is the supremum of $[\mathbf{c}_n, \mathbf{x}]_n$ for $||\mathbf{x}||_n^2 \leqq 1$, $\sigma_n(g) = ||\mathbf{c}_n||_n$. Note that

$$|[\mathbf{c}_n, \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n - \mathbf{Z}_n]_n| \leqq \sigma_n(g)||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n - \mathbf{Z}_n||_n(A) .$$

Since $d_n f_n^2 \to 0$, Theorem 6 implies that

$$[\hat{g}_n - g(\boldsymbol{\theta})]/\sigma_n(g) - [\mathbf{c}_n, \mathbf{Z}_n]_n/||\mathbf{c}||_n \to_P 0.$$

To complete the proof, it suffices to show that

$$W_n = [\mathbf{c}_n, \mathbf{Z}_n]_n/||\mathbf{c}||_n = \sum_{i\in I(n)} (c_{in}, Y_i - E_i)/||\mathbf{c}_n||_n \to_{\mathscr{D}} N(0, 1) .$$

To do so, note that W_n has a moment generating function Ψ_n with logarithm

$$\begin{aligned} \log \Psi_n(t) &= \textstyle\sum_{i\in I(n)} [\log a(\theta_i) - \log a(\theta_i + tc'_{in}) - t(c'_{in}, E_i)] \\ &= \tfrac{1}{2}t^2 \textstyle\sum_{i\in I(n)} (c'_{in}, \Sigma(\theta_i + t'c'_{in})c'_{in}),\ |t|\,||\mathbf{c}_n'||_n(A) \leqq f_n . \end{aligned} \tag{3.17}$$

Here $t' = \alpha t$ for some α, $0 < \alpha < 1$, and $c'_{in} = c_{in}/||\mathbf{c}_n||_n$ for $i \in I(n)$. The equation follows from a standard Taylor expansion, given the observation that $\log a(\cdot)$ has differential $-E(\theta)$ at $\theta \in \Theta^\circ$ and $E(\cdot)$ has differential $\Sigma(\theta)$ at $\theta \in \Theta^\circ$ (see Berk (1972)).

The norm $||\mathbf{c}_n'||_n(A) = 1$. Clearly $||\mathbf{c}_n'||_n(A) \leqq ||\mathbf{c}_n'||_n = 1$. Note that if $\rho_n \mathbf{x} = \mathbf{c}_n'$ and $\mathbf{x} \in \Omega_\infty$, then $|g(\mathbf{x})| = [\mathbf{c}_n, \mathbf{c}_n']_n = \sigma_n(g)$. Thus $||\mathbf{c}_n'||_n(A) = 1$. Since $f_n \to \infty$, Condition 1 implies that

$$\log \Psi_n(t) \to \tfrac{1}{2}t^2 , \qquad |t| \leqq e .$$

By Curtiss (1942), $W_n \to_{\mathscr{D}} N(0, 1)$. □

More generally, the following theorem is available.

THEOREM 11. *Assume Condition* 1 *holds. Let* $g \in \Omega_\infty^*$ *and* $n(g) < \infty$. *Assume* $g(\mathbf{x}) \neq 0$ *for some* $\mathbf{x} \in \Omega_\infty$. *If*

$$[\tau_n(g, A)/\sigma_n(g)]d_n f_n^2 \to 0 ,$$

then

$$[\hat{g}_n - g(\boldsymbol{\theta})]/\sigma_n(g) \to_{\mathscr{D}} N(0, 1) .$$

REMARK. As in the proof of Theorem 8, $\tau_n(g, A)$ is the supremum of $|g(\mathbf{y})|$ for $\mathbf{y} \in \Omega_\infty$ such that $||\rho_n \mathbf{y}||_n(A) \leqq 1$. Note that $\tau_n(g, A) \geqq \tau_n(g, \Omega_\infty^*) = \sigma_n(g)$ and $\sigma_n(g, A) = \tau_n(g)$ for $g \in A$. Thus Theorem 11 is a generalization of Theorem 10.

PROOF. The proof is essentially the same as for Theorem 10. Let $\mathbf{c}_n$, W_n and Ψ_n be defined as in the proof of Theorem 10. Note that

$$|[\mathbf{c}_n, \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n - \mathbf{Z}_n]|_n \leqq \tau_n(g, A)||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n - Z_n||_n(A)$$

and

$$\tau_n(g, A)||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n - \mathbf{Z}_n||_n(A)/\sigma_n(g) \to_P 0 .$$

Thus

$$[\hat{g}_n - g(\theta)]/\sigma_n(g) - W_n \to_P 0 .$$

Equation (3.17) holds for $t' = \alpha t$, $0 < \alpha < 1$, and $c'_{in} = c_{in}/\sigma_n(g)$ for $i \in I(n)$. Since $||\mathbf{c}_n'||_n(A) \leqq 1$, if $t \leqq f_n$, then

$$\begin{aligned} |\textstyle\sum_{i\in I(n)} (c'_{in}, \Sigma(\theta_i + t'c'_{in})c'_{in}) - \sum_{i\in I(n)} (c'_{in}, \Sigma_i c'_{in})| \\ &= |\textstyle\sum_{i\in I(n)} (c'_{in}, \Sigma(\theta_i + t'c'_{in})c'_{in}) - 1| \\ &\leqq \frac{\tau_n(g, A)}{\sigma_n(g)} ||\mathbf{U}_n(\boldsymbol{\theta}_n + t'\mathbf{c}_n', \mathbf{c}_n')||_n(A) \\ &\leqq |t| d_n \tau_n(g, A)/\sigma_n(g) . \end{aligned}$$

For $|t| \leqq e$, $\log \Psi_n(t) \to \frac{1}{2}t^2$ as $n \to \infty$. Thus $W_n \to_{\mathscr{D}} N(0, 1)$. □

3.3. *Verification of regularity conditions when $A = \Omega_\infty^*$.* The conditions used in Theorems 5 through 11 are most readily verified if $A = \Omega_\infty^*$ and $\Omega_\infty \neq \{\mathbf{0}\}$. In this subsection, it is shown that Condition 2 implies Condition 1, and Condition 3 implies Condition 2. These new conditions are defined as follows. In these conditions, $a_n = \dim \Omega_n$ and b_n is the smallest nonnegative number such that $||x_i|| \leqq b_n||\mathbf{x}||_n$ for all $i \in I(n)$ and $\mathbf{x} \in \Omega_n$.

CONDITION 2. There exist constants $e > 0$, $n' \geqq 1$, $d_n \geqq 0$, $n \geqq n'$, and $f_n \geqq e$, $n \geqq n'$, with the following properties:

(a) If $y \in \Omega_n$ and $||\mathbf{y} - \boldsymbol{\theta}_n||_n \leqq f_n$, then $y_i \in \Theta^\circ$, $i \in I(n)$, and

$$\sum_{i \in I(n)} (z_i, [\Sigma(y_i) - \Sigma_i]z_i) \leqq d_n||\mathbf{z}||_n^2||\mathbf{y} - \boldsymbol{\theta}_n||_n\,, \qquad \mathbf{z} \in \Omega_n\,.$$

(b) As $n \to \infty$, $a_n/f_n^2 \to 0$ and $d_n f_n \to 0$.

CONDITION 3. As $n \to \infty$, $a_n b_n^2 \to 0$. In addition, one of the following statements holds:

(a) The convex support C of νY^{-1} is bounded.

(b) For all $i \in J$, $\theta_i \in B$, where B is a compact subset of Θ°.

(c) For some $\varepsilon > 0$ and $\kappa \geqq 0$, if $x, y \in V$, $i \in J$, and $||y - \theta_i|| \leqq \varepsilon$, then $y \in \Theta^\circ$ and

$$(x, [\Sigma(y) - \Sigma_i]x) \leqq \kappa(x, \Sigma_i x)||y - \theta_i||\,.$$

The lemmas which follow may be used to verify that Condition 3 implies Condition 2, and Condition 2 implies Condition 1.

LEMMA 1. *For $n \geqq 1$, $E\{||\mathbf{Z}_n||_n^2\} = a_n$. If Condition 2 holds and if $a_n \to a < \infty$, then $||\mathbf{Z}_n||_n^2 \to_{\mathscr{D}} \chi_a^2$.*

PROOF. There exist $\mathbf{c}_{jn} \in \Omega_n$, $1 \leqq j \leqq a_n$, such that

$$[\mathbf{c}_{jn}, \mathbf{c}_{kn}]_n = 1\,, \quad j = k$$
$$= 0\,, \quad j \neq k\,.$$

The $\mathbf{c}_{jn}$, $1 \leqq j \leqq a_n$, form an orthonormal basis of Ω_n with respect to $[\cdot, \cdot]_n$. By (3.12),

$$W_{jn} = [c_{jn}, \mathbf{Z}_n]_n = \sum_{i \in I(n)} (c_{ijn}, Y_i - E_i)\,.$$

Thus

$$||\mathbf{Z}_n||_n^2 = \sum_{j=1}^{a_n} W_{jn}^2\,.$$

Since $E\{W_{jn}\} = 0$ for $1 \leqq j \leqq a_n$, one has

$$E\{W_{jn}^2\} = \mathrm{Var}\,\{W_{jn}\} = \sum_{i \in I(n)} \mathrm{Var}\,\{(c_{ijn}, Y_i - E_i)\}$$
$$= \sum_{i \in I(n)} (c_{ijn}, \Sigma_i c_{ijn}) = 1\,, \quad 1 \leqq j \leqq a_n\,.$$

Thus $E\{||\mathbf{Z}_n||_n^2\} = a_n$. If $a_n = a$ for n sufficiently large, then similar arguments to those in Theorem 10 may be used to show that

$$\sum_{j=1}^{a} \gamma_j W_{jn} \to_{\mathscr{D}} N(0, \sum_{j=1}^{a} \gamma_j^2)\,, \qquad \boldsymbol{\gamma} \in R^a\,.$$

Thus $\mathbf{W}_n = \{W_{jn}\colon 1 \leqq j \leqq a\}$ converges in distribution to the multivariate normal distribution with zero mean with the identity covariance operator. It follows that $||\mathbf{Z}_n||_n^2 \to_{\mathscr{D}} \chi_a^2$. □

LEMMA 2. *If* $a_n/f_n^2 \to 0$, *then* $\Pr\{||\mathbf{Z}_n||_n < \frac{1}{2}f_n\} \to 1$.

PROOF. Just note that

$$\Pr\{||\mathbf{Z}_n||_n > \tfrac{1}{2}f_n\} \leqq E\{||\mathbf{Z}_n||_n^2\}/(\tfrac{1}{2}f_n)^2 = 4a_n/f_n^2 . \qquad \square$$

LEMMA 3. *Assume that if* $||\mathbf{y} - \boldsymbol{\theta}_n||_n \leqq f_n$ *and* $\mathbf{y} \in \Omega_n$, *then* $y_i \in \Theta^\circ$ *for* $i \in I(n)$. *If*

$$|\textstyle\sum_{i\in I(n)} (z_i, [\Sigma(y_i) - \Sigma_i]z_i)| \leqq d_n||\mathbf{z}||_n^2||\mathbf{y} - \boldsymbol{\theta}_n||_n ,$$
$$\mathbf{y}, \mathbf{z} \in \Omega_n,\ ||\mathbf{y} - \boldsymbol{\theta}_n||_n \leqq f_n ,$$

then

$$||\mathbf{U}_n(\mathbf{y}, \mathbf{z})||_n \leqq d_n||\mathbf{z}||_n||\mathbf{y} - \boldsymbol{\theta}_n||_n , \qquad \mathbf{y}, \mathbf{z} \in \Omega_n,\ ||\mathbf{y} - \boldsymbol{\theta}_n||_n \leqq f_n .$$

PROOF. Note that $||\mathbf{U}_n(\mathbf{y}, \mathbf{z})||_n$ is the smallest nonnegative number such that

$$|\textstyle\sum_{i\in I(n)} (x_i, [\Sigma(y_i) - \Sigma_i]z_i)| \leqq ||\mathbf{U}_n(\mathbf{y}, \mathbf{z})||_n||\mathbf{x}||_n , \qquad \mathbf{x} \in \Omega_n .$$

If $\Omega_n = \{\mathbf{0}\}$, then the lemma is trivially true. If $\Omega_n \neq \{\mathbf{0}\}$, then for $||\mathbf{x}||_n \leqq 1$ and $||\mathbf{z}||_n \leqq 1$, $\mathbf{x}, \mathbf{z} \in \Omega_n$,

$$|\textstyle\sum_{i\in I(n)} (x_i, [\Sigma(y_i) - \Sigma_i]z_i)|$$

does not exceed the supremum of

$$|\textstyle\sum_{i\in I(n)} (z_i, [\Sigma(y_i) - \Sigma_i]z_i)|$$

for $||\mathbf{z}||_n \leqq 1$, $\mathbf{z} \in \Omega_n$. Given this observation, the lemma follows. □

LEMMA 4. *Condition* 2 *implies Condition* 1.

PROOF. This result follows from Lemma 2 and Lemma 3. □

LEMMA 5. *Condition* 3 *implies Condition* 2.

PROOF. Let $\{f_n : n \geqq n'\}$ be an increasing sequence of positive numbers such that $b_n f_n \to 0$ and $a_n/f_n^2 \to 0$. Since $b_n = 0$ if and only if $a_n = 0$, such a sequence can be found.

Assume statement (c) of Condition 3 is true. Let n be large enough so that $b_n f_n \leqq \varepsilon$. Let $\mathbf{y} \in \Omega_n$ and $||y - \boldsymbol{\theta}_n||_n \leqq f_n$. Then $||y_i - \theta_i|| \leqq b_n f_n \leqq \varepsilon$ for $i \in I(n)$. Therefore, $y_i \in \Theta^\circ$ for $i \in I(n)$ and

$$\textstyle\sum_{i\in I(n)} (z_i, [\Sigma(y_i) - \Sigma_i]z_i) \leqq \kappa b_n[\sum_{i\in I(n)} (z_i, \Sigma_i z_i)]||\mathbf{y} - \boldsymbol{\theta}_n||_n , \qquad \mathbf{z} \in \Omega_n .$$

Thus Condition 2 holds with $e = f_1$ and $d_n = \kappa b_n$.

Statement (b) implies statement (c), for $\Sigma(\cdot)$ is continuously differentiable on Θ° and $\Sigma(\theta)$ is positive definite for $\theta \in \Theta^\circ$. Thus Condition 2 holds whenever statement (b) is true.

Statement (a) also implies statement (c). To verify this claim, let $\varepsilon > 0$ and let

$$w = \sup_{c\in C} ||c|| .$$

Let $i \in J$, $x, y \in V$, and $||y - \theta_i|| \leqq \varepsilon$. Since $\Theta^\circ = V$, $y \in \Theta^\circ$. Note that

$$\begin{aligned}(x, \Sigma(y)x) &= \int (x, Y(s) - E(y_i))^2 p(y, s)\, d\nu(s) \\ &\leqq \int (x, Y(s) - E_i)^2 p(y, s)\, d\nu(s) \\ &= \frac{\int (x, Y(s) - E_i)^2 \exp(y - \theta_i, Y(s)) p(\theta_i, s)\, d\nu(s)}{\int \exp(y - \theta_i, Y(s)) p(\theta_i, s)\, d\nu(s)} \\ &\leqq \exp(2w||y - \theta_i||)(x, \Sigma_i x)\,.\end{aligned}$$

Similarly,

$$(x, \Sigma_i x) \leqq \exp(2w||y - \theta_i||)(x, \Sigma(y)x)\,.$$

For some $\kappa \geqq 0$,

$$\exp(2wz) - 1 \leqq \kappa z\,, \qquad 0 \leqq z \leqq \varepsilon\,.$$

Thus

$$(x, [\Sigma(y_i) - \Sigma_i]x) \leqq \kappa(x, \Sigma_i x)||y - \theta_i||$$

if $||y - \theta_i|| \leqq \varepsilon$. □

Given these lemmas, the following corollaries to Theorems 5 through 11 are available.

COROLLARY 1. *Assume that Condition 2 holds. As $n \to \infty$, the probability approaches 1 that $\hat{\boldsymbol{\theta}}_n$ exists.*

COROLLARY 2. *Assume Condition 2 holds. As $n \to \infty$, the probability approaches 1 that*

$$|||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n||_n - ||\mathbf{Z}_n||_n| \leqq d_n||\mathbf{Z}_n||_n^2$$

and

$$||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_n - \mathbf{Z}_n||_n \leqq d_n||\mathbf{Z}_n||_n^2\,.$$

If $a_n \to a < \infty$, then

$$||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n||_n^2 \to_{\mathscr{D}} \chi_a^2\,.$$

COROLLARY 3. *Assume Condition 3 holds. Let $g \in \Omega_\infty{}^*$ and $n(g) < \infty$. Then $\hat{g}_n \to_P g(\boldsymbol{\theta})$ as $n \to \infty$.*

COROLLARY 4. *Assume Condition 2 holds. Let $g \in \Omega_\infty{}^*$, $n(g) < \infty$, and $g(x) \neq 0$ for some $x \in \Omega_\infty$. Then*

$$\hat{\sigma}_n(g)/\sigma_n(g) \to_P 1\,.$$

COROLLARY 5. *Assume Condition 2 holds. Let $0 < \alpha < 1$, and let $\chi^2_{a,\alpha}$ be the upper α-point of the $\chi_a{}^2$ distribution. Assume that $a_n \to a < \infty$. Then as $n \to \infty$,*

$$\Pr\{|\hat{g}_n - g(\boldsymbol{\theta})| \leqq \chi_{a,\alpha}\hat{\sigma}_n(g) \;\; \forall\, g \in \Omega^*_{n\infty}\} \to 1 - \alpha\,.$$

REMARK. This corollary permits construction of simultaneous confidence intervals for all linear functionals $g(\boldsymbol{\theta})$, $g \in \Omega^*_{n\infty}$.

PROOF. Note that $|\hat{g}_n - g(\boldsymbol{\theta})| \leqq \chi_{a,\alpha}\hat{\sigma}_n(g)$ for all $g \in \Omega^*_{n\infty}$ if and only if

$$|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n|_n^2 \leqq \chi^2_{a,\alpha}\,.$$

Condition 2 and Corollary 2 imply that

$$|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n|_n^2 - ||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n||_n^2 \to_P 0$$

and

$$||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n||_n^2 \to_{\mathscr{D}} \chi_a^2 .$$

Thus

$$|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n|_n^2 \to_{\mathscr{D}} \chi_a^2 .$$

The conclusion of the corollary follows immdiately. □

COROLLARY 6. *Assume Condition* 2 *holds and* $d_n f_n^2 \to 0$. *Let* $g \in \Omega_\infty{}^*$ *and* $n(g) < \infty$. *Assume* $g(\mathbf{x}) \neq 0$ *for some* $\mathbf{x} \in \Omega_\infty{}^*$. *Then*

$$[\hat{g}_n - g(\boldsymbol{\theta})]/\sigma_n(g) \to_{\mathscr{D}} N(0, 1) .$$

COROLLARY 7. *Assume Condition* 3 *holds and* $a_n b_n \to 0$. *Let* $g \in \Omega_\infty{}^*$ *and* $n(g) < \infty$. *Assume* $g(\mathbf{x}) \neq 0$ *for some* $\mathbf{x} \in \Omega_\infty$. *Then*

$$[\hat{g}_n - g(\boldsymbol{\theta})]/\sigma_n(g) \to_{\mathscr{D}} N(0, 1) .$$

PROOF. Note that in the proof of Lemma 5, if $a_n b_n \to 0$, then d_n and f_n, $n \geqq 1$, may be chosen so that $b_n f_n^2 \to 0$, $a_n/f_n^2 \to 0$, and $d_n = \kappa b_n$. □

To illustrate use of these corollaries, consider the following example.

EXAMPLE 2. *The Dempster model* (*continued*). In this example, it is often relatively easy to verify Condition 3. If V has dimension v, then $a_n \leqq vp(n)$, with equality if the vectors $\{x_{ij} : 1 \leqq i \leqq n\}$, $1 \leqq j \leqq p(n)$, are independent.

To find b_n, let $\langle \cdot, \cdot \rangle_{p(n)}$ be the inner product of $V^{p(n)}$ such that for $y_j \in V$ and $z_j \in V$, $1 \leqq j \leqq p(n)$,

$$\langle \mathbf{y}, \mathbf{z} \rangle_{p(n)} = \sum_{j=1}^{p(n)} (y_j, z_j) .$$

Let D_n be the linear transformation on $V^{p(n)}$ such that for $\mathbf{y}, \mathbf{z} \in V^{p(n)}$,

$$\langle \mathbf{y}, D_n \mathbf{z} \rangle_{p(n)} = \sum_{i=1}^{n} \sum_{j=1}^{p(n)} \sum_{k=1}^{p(n)} (y_j, \Sigma_i z_k) x_{ij} x_{ik} .$$

Thus if

$$W_{jn} = \sum_{i \in I(n)} x_{ij} Y_i , \qquad 1 \leqq j \leqq p(n) ,$$

and $\mathbf{W}_n = \{W_{jn} : 1 \leqq j \leqq p(n)\}$, then D_n is the covariance operator of $\mathbf{W}_n$. Assume that for some $n' \geqq 1$, D_n is positive definite for $n \geqq n'$.

By Rao (1973, page 60), if $\mathbf{y} \in V^{p(n)}$ and $n \geqq n'$, then the supremum of $\langle \mathbf{y}, \mathbf{z} \rangle_{p(n)}$ for $\mathbf{z} \in V^{p(n)}$, $\langle \mathbf{z}, D_n \mathbf{z} \rangle_{p(n)} \leqq 1$, is $\langle \mathbf{y}, D_n^{-1} \mathbf{y} \rangle_{p(n)}$. Since

$$||d||_n^2 = \langle \mathbf{z}, D_n \mathbf{z} \rangle_{p(n)}$$

if

$$d_i = \sum_{j=1}^{p(n)} x_{ij} z_j , \qquad 1 \leqq i \leqq n ,$$

it follows that b_n is the supremum of $\langle \mathbf{y}, D_n^{-1} \mathbf{y} \rangle_{p(n)}^{\frac{1}{2}}$ for $\mathbf{y} \in V^{p(n)}$ such that for some $i \in I(n)$ and $c \in V$, $||c|| = 1$ and

$$y_j = x_{ij} c , \qquad 1 \leqq j \leqq p(n) .$$

To find a bound for b_n, let γ_n be the smallest eigenvalue of D_n, and let

$$\delta_n = \max_{i \in I(n)} \sum_{j=1}^{p(i)} x_{ij}^2 .$$

Then $b_n^2 \leqq \delta_n/\gamma_n$. Thus $a_n b_n^2 \to 0$ if $p(n)\delta_n/\gamma_n \to 0$, and $a_n b_n \to 0$ if $[p(n)]^2\delta_n/\gamma_n \to 0$.

If

$$g(\boldsymbol{\theta}) = \sum_{j=1}^{p(n)} (c_j, \beta_j)$$

for some $n \geqq n'$, then

$$\sigma_n(g) = \langle \mathbf{c}, D_n^{-1}\mathbf{c}\rangle_{p(n)}$$

and

$$\hat{\sigma}_n(g) = \langle \mathbf{c}, \hat{D}_n^{-1}\mathbf{c}\rangle_{p(n)} ,$$

where

$$\langle \mathbf{y}, \hat{D}_n \mathbf{z}\rangle_{p(n)} = \sum_{i=1}^{n} \sum_{j=1}^{p(n)} \sum_{k=1}^{p(n)} (y_j, \hat{\Sigma}_{in} z_k) x_{ij} x_{ik} , \qquad \mathbf{y}, \mathbf{z} \in V^{p(n)} .$$

For corresponding formulas for $\sigma_n(g)$, see Dempster (1971).

If $p(n)$ is a constant p for n sufficiently large and if a or b holds in Condition 3, then a condition of Haberman (1974, pages 352–373) implies that $a_n b_n \to 0$. In this condition, the distribution of a random vector $\mathbf{X}_n$ is the empirical distribution of the vectors $\mathbf{x}_i = \{x_{ij} : 1 \leqq j \leqq p\}$. As $n \to \infty$, $\mathbf{X}_n \to_{\mathscr{D}} \mathbf{X}$, where

$$E\{\textstyle\sum_{j=1}^{p} d_j X_j\}^2 > 0 , \qquad \mathbf{d} \in R^p,\ \mathbf{d} \neq \mathbf{0} ,$$

and

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{p} x_{ij}^2 \to E\{\textstyle\sum_{j=1}^{p} X_j^2\} .$$

Note that $a_n \to vp$, γ_n/n converges to a positive constant, and $\delta_n/\gamma_n \to 0$.

If $p(n) \to \infty$ and if a or b holds in Condition 3, then $p(n)\delta_n/\gamma_n$ cannot approach 0 if $[p(n)]^2/n$ does not approach 0 as $n \to \infty$. Similarly, $[p(n)]^2\delta_n/\gamma_n$ cannot approach 0 if $[p(n)]^3/n$ does not approach 0. Verification of these claims is straightforward. Let $d < \infty$ be the supremum of the eigenvalues of Σ_i for $i \in J$. If $c \in V$, $||c|| = 1$, $z_j \in R$, $1 \leqq j \leqq p(n)$, and

$$\sum_{j=1}^{p(n)} z_j^2 = 1 ,$$

then

$$\begin{aligned} \gamma_n &\leqq \sum_{j=1}^{p(n)} \sum_{k=1}^{p(n)} \sum_{i=1}^{n} z_j z_k (c, \Sigma_i c) x_{ij} x_{ik} \\ &\leqq d \sum_{j=1}^{p(n)} \sum_{k=1}^{p(n)} z_j z_k \sum_{i=1}^{n} x_{ij} x_{ik} . \end{aligned}$$

The sum of the $p(n)$ eigenvalues of the matrix

$$H_n = \{\textstyle\sum_{i=1}^{n} x_{ij} x_{ik} : 1 \leqq j \leqq p(n),\ 1 \leqq k \leqq p(n)\}$$

is equal to the trace

$$\sum_{i=1}^{n} \sum_{j=1}^{p(n)} x_{ij}^2 \leqq n\delta_n .$$

Thus the smallest eigenvalue of H_n does not exceed $n\delta_n/p(n)$. Therefore,

$$\gamma_n \leqq dn\delta_n/p(n)$$

and

$$p(n)\delta_n/\gamma_n \geqq d^{-1}[p(n)]^2/n .$$

3.4. *Verification of regularity conditions when A_n is finite.* Regularity conditions can sometimes be verified when the set A_n of Condition 1 is finite for $n \geqq 1$. Condition 1 is implied by Condition 4.

CONDITION 4. *A subset $A \subset \Omega_\infty{}^*$ and constants $n' \geqq 1$, $e > 0$, $d_n \geqq 0$, $n \geqq n'$, and $f_n \geqq e$, $n \geqq n'$, exist such that the following conditions hold for $n \geqq n'$:*

(a) *If $\mathbf{x} \in \Omega_\infty$ and $g(\mathbf{x}) = 0$ for $g \in A_n = A \cap \Omega_{n\infty}^*$, then $\rho_n \mathbf{x} = \mathbf{0}$.*
(b) *There are $0 < a_n' < \infty$ elements in A_n, and $f_n{}^2/\log a_n' \to \infty$ as $n \to \infty$.*
(c) *If $\mathbf{y}, \mathbf{z} \in \Omega_n$ and $\|\mathbf{y} - \boldsymbol{\theta}_n\|_n(A) \leqq f_n$, then*

$$y_i \in \Theta^\circ , \quad i \in I(n) ,$$

and

$$\|\mathbf{U}_n(\mathbf{y}, \mathbf{z})\|_n(A) \leqq d_n \|\mathbf{y} - \boldsymbol{\theta}_n\|_n(A) \|\mathbf{z}\|_n(A) .$$

(d) *As $n \to \infty$, $d_n f_n \to 0$.*

To verify this claim, the following lemma is proven.

LEMMA 6. *If Condition 4 is satisfied, then* $\Pr\{\|\mathbf{Z}_n\|_n(A) < \frac{1}{2}f_n\} \to 1$ *as $n \to \infty$.*

PROOF. Let $A_n = \{g_j : 1 \leqq j \leqq a_n'\}$, and let $\mathbf{c}_{jn} \in \Omega_n$ be defined for $1 \leqq j \leqq a_n'$ by the equation

$$g_j(\mathbf{x}) = [\mathbf{c}_{jn}, \rho_n \mathbf{x}]_n , \qquad \mathbf{x} \in \Omega_\infty .$$

Let

$$W_{jn} = \|\mathbf{c}_{jn}\|_n^{-1} \textstyle\sum_{i \in I(n)} (c_{ijn}, Y_i - E_i) , \qquad 1 \leqq j \leqq a_n' ,$$

so that

$$\|\mathbf{Z}_n\|_n(A) = \max_{1 \leqq j \leqq a_n'} |W_{jn}| .$$

Let $\mathbf{c}_{jn}' = c_{jn}/\|\mathbf{c}_{jn}\|$, $1 \leqq j \leqq a_n'$, and let Ψ_{jn} be the moment generating function of W_{jn}.

By the same argument used in (3.17), one finds that

$$\log \Psi_{jn}(t) = \tfrac{1}{2}t^2 \textstyle\sum_{i \in I(n)} (c_{ijn}', \Sigma(\theta_i + t' c_{ijn}') c_{ijn}') , \qquad |t| \leqq f_n ,$$

where $t' = \alpha t$ for some α, $0 < \alpha < 1$. By Bahadur (1971),

$$\Pr\{W_{jn} \geqq \tfrac{1}{2}f_n\} \leqq \exp(-\tfrac{1}{4}f_n{}^2)\Psi_{jn}(\tfrac{1}{2}f_n)$$

and

$$\Pr\{-W_{jn} \geqq \tfrac{1}{2}f_n\} \leqq \exp(-\tfrac{1}{4}f_n{}^2)\Psi_{jn}(-\tfrac{1}{2}f_n)$$

for $1 \leqq j \leqq a_n'$. For $1 \leqq j \leqq a_n'$,

$$\log \Psi_{jn}(\tfrac{1}{2}f_n) \quad \text{and} \quad \log \Psi_{jn}(-\tfrac{1}{2}f_n) \quad \text{do not exceed} \quad \tfrac{1}{8}f_n{}^2(1 + d_n f_n/2) .$$

By the Bonferroni inequality,

$$\begin{aligned} \Pr\{\|\mathbf{Z}_n\|_n(A) \geqq \tfrac{1}{2}f_n\} &\leqq 2a_n' \exp[-\tfrac{1}{8}f_n{}^2(1 - d_n f_n/2)] \\ &= 2\exp[\log a_n' - \tfrac{1}{8}f_n{}^2(1 - d_n f_n/2)] . \end{aligned}$$

Since $\log a_n'/f_n{}^2 \to 0$, the conclusion of the lemma follows. □

To illustrate use of Condition 4, consider the following example.

EXAMPLE 1. *The Rasch model (continued).* In this example, $I(n) = \{(i, j) : 1 \leqq i \leqq r_n, 1 \leqq j \leqq c_n\}$. The sequence $\{r_n : n \geqq 1\}$ is strictly increasing, and

$\{c_n : n \geqq 1\}$ is a nondecreasing sequence such that $c_n \to \infty$ as $n \to \infty$. For convenience, assume that $r_n \geqq c_n$ for $n \geqq 1$. The space Ω_∞ consists of $\mathbf{x} = \{x_{ij} : i \geqq 1, j \geqq 1\}$ such that

$$x_{ij} = y_i + z_j\,, \qquad i \geqq 1, j \geqq 1\,,$$

for some $y_i \in V$, $i \geqq 1$, and $z_j \in V$, $j \geqq 1$. The space Ω_n consists of $\mathbf{x} = \{x_{ij} : 1 \leqq i \leqq r_n, 1 \leqq j \leqq c_n\}$ such that

$$x_{ij} = y_i + z_j\,, \qquad 1 \leqq i \leqq r_n,\ 1 \leqq j \leqq c_n\,,$$

for some $y_i \in V$, $1 \leqq i \leqq r_n$, and $z_j \in V$, $1 \leqq j \leqq c_n$. To simplify a rather complex analysis, assume that for some compact set $T \subset \Theta^\circ$, $\alpha_i + \beta_j \in T$ for $i \geqq 1$ and $j \geqq 1$.

The estimates $\hat{\alpha}_i$ and $\hat{\beta}_j$ have the consistency properties

$$\max_{1 \leqq i \leqq r_n} ||\hat{\alpha}_i - \alpha_i|| \to_P 0$$

and

$$\max_{1 \leqq j \leqq c_n} ||\hat{\beta}_j - \beta_j|| \to_P 0$$

whenever $c_n^{-1} \log r_n \to 0$. If $c_n^{-1}(\log r_n)^2 \to 0$, then nonzero linear combinations such as

$$\textstyle\sum_{i=1}^m (h_i, \hat{\alpha}_i)\,,$$

$h_i \in V$ for $1 \leqq i \leqq m$, are asymptotically normal. If $c_n^{-2} r_n (\log r_n)^2 \to 0$, then nonzero linear combinations such as

$$\textstyle\sum_{j=1}^{m'} (h_j', \hat{\beta}_j)\,,$$

$h_j' \in V$ for $1 \leqq j \leqq m'$, are also asymptotically normal. If

$$\mathbf{\Sigma}_{i+n} = \textstyle\sum_{j=1}^{c_n} \mathbf{\Sigma}_{ij}\,, \qquad 1 \leqq i \leqq r_n\,,$$
$$\mathbf{\Sigma}_{+jn} = \textstyle\sum_{i=1}^{r_n} \mathbf{\Sigma}_{ij}\,, \qquad 1 \leqq j \leqq c_n\,,$$

then the asymptotic variance of

$$\textstyle\sum_{i=1}^m (h_i, \hat{\alpha}_i)$$

can be approximated by

$$\textstyle\sum_{i=1}^m (h_i, \mathbf{\Sigma}_{i+n}^{-1} h_i)\,,$$

and the asymptotic variance of

$$\textstyle\sum_{j=1}^{m'} (h_j', \hat{\beta}_j)\,,$$

can be approximated by

$$\textstyle\sum_{j=1}^{m'} (h_j', \mathbf{\Sigma}_{+jn}^{-1} h_j')\,.$$

These claims are verified through a careful selection of a set $A \subset \Omega_\infty^*$.

To define $A \subset \Omega_\infty^*$, let e_k, $1 \leqq k \leqq v = \dim V$, be a basis of V. Assume that $||e_k|| = 1$ for $1 \leqq k \leqq v$. For $1 \leqq k \leqq v$, $i \geqq 1$, and $j \geqq 1$, let

$$\delta_{ijk}(\mathbf{x}) = (e_k, x_{ij})\,.$$

Let $A = \{\delta_{ijk}: i \geqq 1, j \geqq 1, 1 \leqq k \leqq v\}$, so that $A_n = \{\delta_{ijk}: 1 \leqq i \leqq r_n, 1 \leqq j \leqq c_n, 1 \leqq k \leqq v\}$ and $a_n' = r_n c_n v$. Note that if $\mathbf{x} \in \Omega_\infty$ and $\delta_{ijk}(\mathbf{x}) = 0$, $1 \leqq i \leqq r_n$, $1 \leqq j \leqq c_n$, $1 \leqq k \leqq v$, then $x_{ij} = 0$, $1 \leqq i \leqq r_n$, $1 \leqq j \leqq c_n$.

By (3.10), if $\mathbf{y} \in \Omega_n$, then $\|\mathbf{y}\|_n(A)$ is the maximum value of

$$|(e_k, y_{ij})|/\sigma_n(\delta_{ijk}), \qquad 1 \leqq i \leqq r_n,\ 1 \leqq j \leqq c_n,\ 1 \leqq k \leqq v .$$

To evaluate $\|\mathbf{y}\|_n(A)$, a procedure for computation of $\sigma_n(g)$ is required for $g \in \Omega_n{}^*$. Note that $h_{ijn}(g) \in V$, $1 \leqq i \leqq r_n$, $1 \leqq j \leqq c_n$, may be defined so that

$$g(\mathbf{x}) = \sum_{i=1}^{r_n} \sum_{j=1}^{c_n} (h_{ijn}(g), x_{ij}), \qquad \mathbf{x} \in \Omega_\infty ,$$

and $h_{ijn}(g) = 0$ if neither i nor j is 1. Let

$$\mathbf{d}_n(g) = u_n(g) + w_{in}(g) + z_{jn}(g), \qquad 1 \leqq i \leqq r_n,\ 1 \leqq j \leqq c_n ,$$

where $u_n(g)$, $w_{in}(g)$, $1 \leqq i \leqq r_n$, and $z_{jn}(g)$, $1 \leqq j \leqq c_n$, are in V,

$$\sum_{i=1}^{r_n} \Sigma_{i+n} w_{in}(g) = 0 , \tag{3.18}$$

$$\sum_{j=1}^{c_n} \Sigma_{+jn} z_{jn}(g) = 0 , \tag{3.19}$$

and

$$[\mathbf{d}_n(g), \rho_n(\mathbf{x})]_n = g(\mathbf{x}) , \qquad \mathbf{x} \in \Omega_\infty .$$

Note that $\sigma_n(g) = \|\mathbf{d}_n(g)\|_n$.

If $x_{ij} = y \in V$, $i \geqq 1$, $j \geqq 1$, then

$$[\mathbf{d}_n(g), \rho_n(\mathbf{x})]_n = (u_n(g), \Sigma_{++n} y) = (h_{++n}(g), y) ,$$

where

$$\Sigma_{++n} = \sum_{i=1}^{r_n} \sum_{j=1}^{c_n} \Sigma_{ij}$$

and

$$h_{++n}(g) = \sum_{i=1}^{r_n} \sum_{j=1}^{c_n} h_{ijn}(g) .$$

Thus

$$\Sigma_{++n} u_n(g) = h_{++n}(g) . \tag{3.20}$$

Let

$$h_{i+n}(g) = \sum_{j=1}^{c_n} h_{ijn}(g) , \qquad 1 \leqq i \leqq r_n ,$$
$$h_{+jn}(g) = \sum_{i=1}^{r_n} h_{ijn}(g) , \qquad 1 \leqq j \leqq c_n .$$

Then similar arguments to those used to derive (3.20) may be employed to show that

$$\Sigma_{i+n} w_{in}(g) + \sum_{j=1}^{c_n} \Sigma_{ij} z_{jn}(g) = h_{i+n}(g) - \Sigma_{i+n} \Sigma_{++n}^{-1} h_{++}(g)_n , \qquad 1 \leqq i \leqq r_n , \tag{3.21}$$

$$\Sigma_{+jn} z_{jn}(g) + \sum_{i=1}^{r_n} \Sigma_{ij} w_{in}(g) = h_{+jn}(g) - \Sigma_{+jn} \Sigma_{++n}^{-1} h_{++n}(g) , \qquad 1 \leqq j \leqq c_n . \tag{3.22}$$

Equations (3.18)—(3.22) determine $u_n(g)$, $w_{in}(g)$, $1 \leqq i \leqq r_n$, and $z_{jn}(g)$, $1 \leqq j \leqq c_n$.

Given these coefficients,

$$\begin{aligned}||\mathbf{d}_n(g)||_n^2 &= (u_n(g), \Sigma_{++n}u_n(g)) + \sum_{i=1}^{r_n}(w_{in}(g), \Sigma_{i+n}w_{in}(g)) \\ &\quad + \sum_{j=1}^{c_n}(z_{jn}(g), \Sigma_{+jn}z_{jn}(g)) + 2\sum_{i=1}^{r_n}\sum_{j=1}^{c_n}(w_{in}(g), \Sigma_{ij}z_{jn}(g)) \\ &= (u_n(g), h_{++n}(g)) + \sum_{i=1}^{r_n}(w_{in}(g), h_{i+n}(g)) + \sum_{j=1}^{c_n}(z_{jn}(g), h_{+jn}(g)) .\end{aligned}$$

In practice, approximations for $\sigma_n(g)$ are helpful. To find such an approximation, let $\gamma_n > 0$ be the largest number such that for $1 \leqq i \leqq r_n$, $1 \leqq j \leqq c_n$,

$$(x, \Sigma_{ij}x) \geqq \gamma_n\left[\frac{1}{c_n}(x, \Sigma_{i+}x) + \frac{1}{r_n}(x, \Sigma_{+jn}x)\right], \qquad x \in V .$$

Note that for some $\gamma > 0$, $\gamma_n \geqq \gamma$ for all n.

For $1 \leqq i \leqq r_n$, $1 \leqq j \leqq c_n$, let

$$\begin{aligned} w'_{in}(g) &= \Sigma_{i+n}^{-1}h_{i+n}(g) - \Sigma_{++n}^{-1}h_{++n}(g) , \\ z'_{jn}(g) &= \Sigma_{+jn}^{-1}h_{+jn}(g) - \Sigma_{++n}^{-1}h_{++n}(g) , \\ w''_{in}(g) &= w_{in}(g) - w'_{in}(g) , \\ z''_{jn}(g) &= z_{jn}(g) - z'_{jn}(g) , \\ \Sigma'_{ijn} &= \Sigma_{ij} - \gamma_n\left(\frac{1}{c_n}\Sigma_{i+n} + \frac{1}{r_n}\Sigma_{+jn}\right) , \\ d'_{ijn}(g) &= u_n(g) + w'_{in}(g) + z'_{jn}(g) , \\ d''_{ijn}(g) &= w''_{in}(g) + z''_{jn}(g) = d_{ijn}(g) - d'_{ijn}(g) .\end{aligned}$$

Note that

$$||\mathbf{d}_n'(g)||_n - ||\mathbf{d}_n''(g)||_n \leqq \sigma_n(g) = ||\mathbf{d}_n(g)||_n \leqq ||\mathbf{d}_n'(g)||_n + ||\mathbf{d}_n''(g)||_n , \tag{3.23}$$

and

$$\begin{aligned}||\mathbf{d}_n'(g)||_n^2 &= \sum_{i=1}^{r_n}(h_{i+n}(g), \Sigma_{i+n}^{-1}h_{i+n}(g)) + \sum_{j=1}^{c_n}(h_{+jn}(g), \Sigma_{+jn}^{-1}h_{+jn}(g)) \\ &\quad - 3(h_{++n}(g), \Sigma_{++n}^{-1}h_{++n}(g)) \\ &\quad + 2\sum_{i=1}^{r_n}\sum_{j=1}^{c_n}(h_{i+n}(g), \Sigma_{i+n}^{-1}\Sigma_{ijn}\Sigma_{+jn}^{-1}h_{+jn}(g)) .\end{aligned} \tag{3.24}$$

Let

$$\begin{aligned} l_{in}(g) &= -\sum_{j=1}^{c_n}\Sigma_{ijn}\Sigma_{+jn}^{-1}h_{+jn}(g) + \Sigma_{i+n}\Sigma_{++n}^{-1}h_{++n}(g) , && 1 \leqq i \leqq r_n , \\ m_{jn}(g) &= -\sum_{i=1}^{r_n}\Sigma_{ijn}\Sigma_{i+n}^{-1}h_{i+n}(g) + \Sigma_{+jn}\Sigma_{++n}^{-1}h_{++n}(g) , && 1 \leqq j \leqq c_n .\end{aligned}$$

Then

$$\begin{aligned} \Sigma_{i+n}w''_{in}(g) + \sum_{j=1}^{c_n}\Sigma_{ijn}z''_{jn}(g) &= l_{in}(g) , && 1 \leqq i \leqq r_n , \\ \Sigma_{+jn}z''_{jn}(g) + \sum_{i=1}^{r_n}\Sigma_{ijn}w''_{in}(g) &= m_{jn}(g) , && 1 \leqq j \leqq c_n .\end{aligned}$$

By Rao (1973, page 60),

$$||\mathbf{d}_n''(g)||_n^2 = \sum_{i=1}^{r_n}(w''_{in}(g), l_{in}(g)) + \sum_{j=1}^{c_n}(z''_{jn}(g), m_{jn}(g))$$

is the largest value of

$$[\sum_{i=1}^{r_n}(x_i, l_{in}(g)) + \sum_{j=1}^{c_n}(y_j, m_{jn}(g))]^2$$

for $x_i \in V$, $1 \leqq i \leqq r_n$, and $y_j \in V$, $1 \leqq j \leqq c_n$, such that

$$\sum_{i=1}^{r_n} \Sigma_{i+n} x_i = \sum_{j=1}^{c_n} \Sigma_{+jn} y_j = 0$$

and

$$\begin{aligned} D(\mathbf{x}, \mathbf{y}) = & \sum_{i=1}^{r_n} (x_i, \Sigma_{i+n} x_i) + \sum_{j=1}^{c_n} (y_j, \Sigma_{+jn} y_j) \\ & + \sum_{i=1}^{r_n} \sum_{j=1}^{c_n} (x_i, \Sigma_{ijn} y_j) + \sum_{i=1}^{r_n} \sum_{j=1}^{c_n} (y_j, \Sigma_{ijn} x_i) \\ \leqq & \; 1 . \end{aligned}$$

Since

$$\begin{aligned} D(\mathbf{x}, \mathbf{y}) = & \gamma_n \left[\sum_{i=1}^{r_n} \left(x_i, \left(\Sigma_{i+n} + \frac{1}{r_n} \Sigma_{++n} \right) x_i \right) + \sum_{j=1}^{c_n} \left(y_j, \left(\Sigma_{+jn} + \frac{1}{c_n} \Sigma_{++n} \right) y_j \right) \right] \\ & + \sum_{i=1}^{r_n} \sum_{j=1}^{c_n} (x_i + y_j, \Sigma'_{ijn}(x_i + y_j)) \\ \geqq & \; \gamma_n \left[\sum_{i=1}^{r_n} (x_i, \Sigma_{i+n} x_i) + \sum_{j=1}^{c_n} (y_j, \Sigma_{+jn} y_j) \right] , \end{aligned}$$

it follows that

$$(3.25) \qquad \begin{aligned} \|\mathbf{d}_n''(g)\|_n \leqq & \gamma_n^{-1} \left[\sum_{i=1}^{r_n} (l_{in}(g), \Sigma_{i+n}^{-1} l_{in}(g)) \right. \\ & \left. + \sum_{j=1}^{c_n} (m_{jn}(g), \Sigma_{+jn}^{-1} m_{jn}(g)) \right] . \end{aligned}$$

For $1 \leqq i \leqq r_n$, $1 \leqq j \leqq c_n$, and $1 \leqq k \leqq v$, one has

$$\begin{aligned} \|\mathbf{d}_n'(\delta_{ijk})\|_n^2 = & (e_k, \Sigma_{i+n}^{-1} e_k) + (e_k, \Sigma_{+jn}^{-1} e_k) - 3(e_k, \Sigma_{++n}^{-1} e_k) \\ & + 2(e_k, \Sigma_{i+n}^{-1} \Sigma_{ijn} \Sigma_{+jn}^{-1} e_k) \end{aligned}$$

and

$$\begin{aligned} \|\mathbf{d}_n''(\delta_{ijk})\|_n^2 \leqq & \gamma_n^{-1} \left[\sum_{i'=1}^{r_n} (e_k, \Sigma_{+jn}^{-1} \Sigma_{i'jn} \Sigma_{i'+n}^{-1} \Sigma_{i'jn} \Sigma_{+jn}^{-1} e_k) \right. \\ & \left. + \sum_{j'=1}^{c_n} (e_k, \Sigma_{i+n}^{-1} \Sigma_{ij'n} \Sigma_{+j'n}^{-1} \Sigma_{ij'n} \Sigma_{i+n}^{-1} e_k) - 2(e_k, \Sigma_{++n}^{-1} e_k) \right] . \end{aligned}$$

For some $\tau \geqq 0$,

$$\begin{aligned} & |\sigma_n(\delta_{ijk}) - [(e_k, \Sigma_{i+n}^{-1} e_k) + (e_k, \Sigma_{+jn}^{-1} e_k) - (e_k, \Sigma_{++n}^{-1} e_k)]^{\frac{1}{2}}| \\ & \qquad \leqq \tau/(r_n c_n)^{\frac{1}{2}} , \qquad 1 \leqq i \leqq r_n,\ 1 \leqq j \leqq c_n,\ 1 \leqq k \leqq v,\ n \geqq 1 . \end{aligned}$$

If as in the Poisson distribution, $\Sigma_{ij} = \Sigma_{i+n} \Sigma_{++n}^{-1} \Sigma_{+jn}$, then τ may be set equal to 0.

Let

$$0 < \tau_1 < (x, \Sigma_{ij} x) < \tau_2$$

for $i \geqq 1$, $j \geqq 1$, and $x \in V$ such that $\|x\| = 1$. For some $n' \geqq 1$, if $n \geqq n'$, then $\tau_2^{-1}(r_n^{-1} + c_n^{-1}) < \sigma_n^2(\delta_{ijk}) < \tau_1^{-1}(r_n^{-1} + c_n^{-1})$, $1 \leqq i \leqq r_n$, $1 \leqq j \leqq c_n$, $1 \leqq k \leqq v$, and

$$\tau_2^{-1}(r_n^{-1} + c_n^{-1}) < b_n^2 < \tau_1^{-1}(r_n^{-1} + c_n^{-1}) .$$

Note that this inequality and the fact that $a_n = (r_n + c_n - 1)$ imply that Condition 3 of Section 3.3 cannot hold. Nonetheless, b_n does approach 0.

On the other hand, Condition 4 may be verified under the assumption that $c_n^{-1} \log r_n \to 0$. Let f_n be chosen so that $f_n^2/\log(r_n c_n v) \to \infty$ and $f_n^2(r_n^{-1} + c_n^{-1}) \to 0$. As in Condition 3, define $\varepsilon > 0$ and $d \geqq 0$ so that for $x, y \in V$, $i \geqq 1$, $j \geqq 1$, and $||y - \alpha_i - \beta_j|| \leqq \varepsilon$, one has $y \in \Theta^\circ$ and

$$(x, [\Sigma(y) - \Sigma_{ij}]x) \leqq \kappa(x, \Sigma_{ij}x)||y - \alpha_i - \beta_j|| .$$

Note that for some $\tau_3 > 0$,

$$||y_{ij}|| \leqq \tau_3||\mathbf{y}||_n(A)(r_n^{-1} + c_n^{-1})^{\frac{1}{2}} , \qquad \mathbf{y} \in \Omega_n .$$

Let n' be sufficiently large so that

$$\tau_3 f_n(r_n^{-1} + c_n^{-1})^{\frac{1}{2}} \leqq \varepsilon , \qquad n \geqq n' .$$

If $\mathbf{y} \in \Omega_n$ and $||\mathbf{y} - \boldsymbol{\theta}_n||_n(A) \leqq f_n$, then $y_{ij} \in \Theta^\circ$ for $1 \leqq i \leqq r_n$, $1 \leqq j \leqq c_n$.

To bound $||\mathbf{U}_n(\mathbf{y}, \mathbf{z})||_n(A)$ for $n \geqq n'$, $\mathbf{y}, \mathbf{z} \in \Omega_n$, and $||\mathbf{y} - \boldsymbol{\theta}_n||_n(A) \leqq f_n$, note that for $1 \leqq i \leqq r_n$, $1 \leqq j \leqq c_n$, and $1 \leqq k \leqq v$,

$$\begin{aligned} &|(e_k, U_{ijn}(\mathbf{y}, \mathbf{z}))| \\ &\quad = |\textstyle\sum_{i'=1}^{r_n} \sum_{j'=1}^{c_n} (d_{i'j'n}(\delta_{ijk}), [\Sigma(y_{i'j'}) - \Sigma_{i'j'}]z_{i'j'})| \\ &\quad \leqq \kappa\tau_3^2(r_n^{-1} + c_n^{-1})[||\mathbf{y} - \boldsymbol{\theta}_n||_n(A)][||\mathbf{z}||_n(A)] \textstyle\sum_{i'=1}^{r_n} \sum_{j'=1}^{c_n} ||d_{i'j'n}(\delta_{ijk})|| . \end{aligned}$$

Note that for some $\tau' \geqq 0$,

$$\begin{aligned} \textstyle\sum_{i'=1}^{r_n} \sum_{j'=1}^{c_n} ||d_{i'j'n}(\delta_{ijk})|| &\leqq \textstyle\sum_{i'=1}^{r_n} \sum_{j'=1}^{c_n} ||d'_{i'j'n}(\delta_{ijk})|| + \sum_{i'=1}^{r_n} \sum_{j'=1}^{c_n} ||d''_{i'j'n}(\delta_{ijk})|| , \\ \textstyle\sum_{i'=1}^{r_n} \sum_{j'=1}^{c_n} ||d'_{i'j'n}(\delta_{ijk})|| &= ||\Sigma_{i+n}^{-1} e_k + \Sigma_{+jn}^{-1} e_k - \Sigma_{++n}^{-1} e_k|| + (r_n - 1)||\Sigma_{+jn}^{-1} e_k|| \\ &\quad + (c_n - 1)||\Sigma_{i+n}^{-1} e_k|| + (r_n - 1)(c_n - 1)||\Sigma_{++n}^{-1} e_k|| \\ &\leqq 3\tau_1^{-1} , \end{aligned}$$

and

$$\begin{aligned} \textstyle\sum_{i'=1}^{r_n} \sum_{j'=1}^{c_n} ||d''_{i'j'n}(\delta_{ijk})|| &\leqq (r_n c_n)^{\frac{1}{2}}||\mathbf{d}_n''(\delta_{ijk})||_n \\ &\leqq \tau' . \end{aligned}$$

Thus parts (c) and (d) of Condition 4 are satisfied by

$$d_n = \tau_2^{\frac{1}{2}}(3\tau_1^{-1} + \tau')\kappa\tau_3^2(r_n^{-1} + c_n^{-1})^{\frac{1}{2}} .$$

It follows from Theorem 5 that the probability approaches 1 that $\hat{\boldsymbol{\theta}}_n$ exists. Given Theorem 6,

$$\max_{1\leqq i\leqq r_n; 1\leqq j\leqq c_n} ||\hat{\theta}_{ijn} - \theta_{ij}|| \to_P 0 .$$

This observation or Theorem 8 implies that $\hat{\alpha}_{in} \to_P \alpha_i$, $i \geqq 1$, and $\hat{\beta}_{jn} \to_P \beta_j$, $j \geqq 1$. Indeed, one has the stronger result

$$\begin{aligned} \max_{1\leqq i\leqq r_n} ||\hat{\alpha}_{in} - \alpha_i|| \to_P 0 , \\ \max_{1\leqq j\leqq c_n} ||\hat{\beta}_{jn} - \beta_j|| \to_P 0 . \end{aligned}$$

These conclusions are also reached by Lord (1975) in the binomial case.

In general, the consistency and existence results cannot be improved. For example, if each Y_{ij} is binomial, then the probability that $\hat{\boldsymbol{\theta}}_n$ does not exist exceeds the probability that $Y_{i+n} = c_n$ for some subject i, $1 \leqq i \leqq r_n$. This probability is

$$1 - \textstyle\prod_{i=1}^{r_n} [1 - \prod_{j=1}^{n} p(\alpha_i + \beta_j, 1)] .$$

This probability approaches 0 if and only if $r_n \exp(-c_n) \to 0$ as $n \to \infty$.

To derive asymptotic normality results, assume that

$$c_n^{-1}(\log r_n)^2 \to 0 .$$

Select f_n so that $f_n^2/\log(r_n c_n v) \to \infty$ and $f_n^4(r_n^{-1} + c_n^{-1}) \to 0$. Let

$$g(\boldsymbol{\theta}) = \sum_{i=1}^{m} (h_i, \alpha_i) + \sum_{j=2}^{m'} (h_j', \beta_j)$$

for some $h_i \in V$, $1 \leqq i \leqq m$, such that some h_i is not 0, and for some $h_j' \in V$, $1 \leqq j \leqq m'$. Then

$$[\hat{g}_n - g(\boldsymbol{\theta})]/\sigma_n(g) \to_{\mathscr{D}} N(0, 1) , \tag{3.26}$$

$$\sigma_n(g) \to 0 , \tag{3.27}$$

and

$$\hat{\sigma}_n(g)/\sigma_n(g) \to_P 1 . \tag{3.28}$$

The last two relationships follow from Theorem 7 and Theorem 9 and the fact that $f_n b_n \to 0$. Note that

$$\tau_n(g, A) \leqq b_n(\sum_{i=1}^{m} ||h_i|| + 2 \sum_{j=2}^{m'} ||h_j'||)$$

and

$$\sigma_n(g) \geqq ||h_i||^2/(h_i, \Sigma_{i+n} h_i)^{\frac{1}{2}}$$
$$\geqq \tau_1^{-\frac{1}{2}}/c_n^{\frac{1}{2}} , \qquad 1 \leqq i \leqq n .$$

Thus $\tau_n(g, A)/\sigma_n(g)$ is bounded above for all n. Asymptotic normality then follows from Theorem 11.

If

$$g(\boldsymbol{\theta}) = \sum_{j=2}^{m'} (h_j', \beta_j)$$

for some $h_j' \in V$, $1 \leqq j \leqq r_n$, such that some $h_j' \neq 0$, then a similar argument may be used to show that (3.27) and (3.28) hold. One may verify (3.26) if $c_n^{-2} r_n (\log r_n)^2 \to 0$. How much this condition may be relaxed is unclear.

To approximate $\sigma_n(g)$ if

$$g(\boldsymbol{\theta}) = \sum_{i=1}^{m} (h_i, \alpha_i) + \sum_{j=1}^{m'} (h_j, \beta_j) ,$$

note that for some $\tau'(g)$,

$$||d_n''(g)||_n \leqq \tau'(g)/(r_n c_n)^{\frac{1}{2}} .$$

Let some h_i or h_j' be nonzero and let

$$h_1' = \sum_{i=1}^{m} h_i - \sum_{j=2}^{m'} h_j' .$$

Then (3.24) implies that

$$[\sigma_n(g)]^{-1}[\sum_{i=1}^{m} (h_i, \Sigma_{i+n}^{-1} h_i) + \sum_{j=1}^{m'} (h_j', \Sigma_{+jn}^{-1} h_j')]^{\frac{1}{2}} \to 1$$

as $n \to \infty$.

REFERENCES

[1] ANDERSEN, A. H. (1974). Multidimensional contingency tables. *Scand. J. Statist.* **1** 115–127.

[2] Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *J. Roy. Statist. Soc. B* **32** 283–301.

[3] Andersen, E. B. (1972). The numerical solution of a set of conditional estimation equations. *J. Roy. Statist. Soc. B* **34** 42–54.

[4] Andersen, E. B. (1973). *Conditional Inference and Models for Measuring.* Mentalhygiejnisk Forskningsinstitut, Copenhagen.

[5] Bahadur, R. R. (1971). *Some Limit Theorem in Statistics. SIAM*, Philadelphia.

[6] Barndorff-Nielsen, O. (1973). *Exponential Families and Conditioning.* Ph.D. thesis, Univ. of Copenhagen.

[7] Berk, R. H. (1972). Consistency and asymptotic normality of MLE's for exponential models. *Ann. Math. Statist.* **43** 193–204.

[8] Bock, R. D. and Lieberman, M. L. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika* **35** 179–197.

[9] Curtiss, J. H. (1942). A note on the theory of moment generating functions. *Ann. Math. Statist.* **13** 430–433.

[10] Dempster, A. P. (1971). An overview of multivariate data analysis. *J. Multivariate Anal.* **1** 316–346.

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

[12] Kantorovich, L. and Akilov, G. P. *Functional Analysis in Normed Spaces* (translated by D. G. Brown). Pergamon Press, Oxford.

[13] Lord, F. (1975). Consistent estimation when number of variables and number of parameters increases without limit (with psychometric applications). Educational Testing Service, Princeton.

[14] Nelder, J. A. and Wedderburn, R. W. N. (1972). Generalized linear models. *J. Roy. Statist. Soc. A* **35** 370–384.

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

[16] Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests.* Danish Institute of Educational Research, Copenhagen.

[17] Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **4** 321–334, Univ. of California Press.

[18] Rockafeller, R. T. (1970). *Convex Analysis.* Princeton Univ. Press.

[19] Wright, B. D. and Douglas, G. A. (1976). Better procedures for sample-free item analysis. Research memorandum no. 20, Statistical Laboratory, Dept. of Education, Univ. of Chicago.

Department of Statistics
University of Chicago
5734 University Avenue
Chicago, Illinois 60637