Sankhyā : The Indian Journal of Statistics
1987, Volume 49, Series B, Pt. 1, pp. 97-102.

ON ZERO CELLS IN LOG-LINEAR MODELS

By RAHUL MUKERJEE
Indian Statistical Institute

SUMMARY. This paper considers the analysis of categorical data under the log-linear model when there are some observed zero cell frequencies. A linear programming formulation is developed for identifying the cells for which the maximum likelihood estimates (MLEs) exist finitely and also the cells for which only 'extended' MLEs exist.

1. Introduction and preliminaries

The analysis of frequency data under log-linear models has attracted considerable attention in recent years (for comprehensive lists of references upto various stages, see Haberman (1974, 1978, 1979) and Bishop, Fienberg and Holland (1975)). In maximum likelihood estimation of the relevant parameters under such models, sometimes problems arise because of observed zero frequencies in some cells. This paper attempts to provide linear programming formulations for handling some of these problems.

To formalize the ideas, attention will be restricted to the Poisson model, but it is well known (vide Birch (1963), Haberman (1974, Ch. 2)) that the results so obtained will cover some other models (e.g. multinomial) as well. Following Harberman (1974, pp. 6–7), in the Poisson model one considers the random vector $n = (n_1, ..., n_q)'$, where $n_1, ..., n_q$ are independent Poisson variates with $E(n_i) = m_i\ (> 0)$, $i = 1, ..., q$. Writing $\mu_i = \log m_i$, $\mu = (\mu_1, ..., \mu_q)'$, it will be assumed that $\mu \in \mathcal{M}$, where $\mathcal{M}$ is a p $(0 < p \leqslant q)$ dimensional linear manifold contained in $\mathcal{R}^q$ the q-dimensional Euclidean space. In such a set-up, if $n_i > 0$ for each i, then the maximum likelihood estimate (MLE) $\hat{\mu}$ of μ exists finitely (Haberman, 1974, pp. 38), while if $n_i = 0$ for some values of i then the MLE may not exist finitely. In fact, in this connexion, the following result holds (Haberman, 1974, pp. 38).

Theorem 1. : *A necessary and sufficient condition that the MLE $\hat{\mu}$ exist finitely is that there does not exist $\mu \in \mathcal{M}$ such that $\mu \neq 0$, $\mu \leqslant 0$ and $n'\mu = 0$.*

If the above condition does not hold Haberman (1974, 402–404) suggests extended MLE for μ as follows. Let $I = \{1, 2, ..., q\}$, $C = \{\mu \in \mathcal{M} : \mu \neq 0,\ \mu \leqslant 0,\ n'\mu = 0\}$. When C is non-empty, define for

---

AMS (1980) *subject classification* : 62F10, 90C05.

Key words and phrases : Linear programming; Maximum likelihood; Zero cell.

B1–13

$\mu \in C$, $J(\mu) = \{i \in I : \mu_i < 0\}$. Let $I^* = \bigcup_{\mu \in C} J(\mu)$, $I_0 = I - I^*$. Then Haberman shows that finite MLE exists for μ_i if $i \in I_0$, while if $i \in I^*$ the (extended) MLE of μ_i turns out to be $-\infty$.

While the above results are theoretically elegant, in practical applications, specially when the number of classes is large, actual verification of the condition of Theorem 1.1 or identification of the sets I_0 and I^* may be troublesome. The present work is concerned mainly with the development of an algorithm for identifying I_0 and I^*. It may be noted that the problem of an algorithmic formulation of Theorem 1.1 is fairly straightforward (in fact this might be already known, although the author is not aware of any references), but, for the sake of completeness, that has also been presented as a passing remark.

2. The algorithm

Adopting a matrix theoretic approach, let $A = (a_1, a_2, \ldots, a_q)$ be a $p \times q$ matrix whose rows form a basis of $\mathcal{M}$. Let $I_1 = \{i \in I : n_i < 0\}$, $I_2 = I - I_1$. Clearly, if I_2 be empty then finite MLE of μ exists. Suppose I_2 is nonempty. Let $A_1(A_2)$ be a submatrix of A consisting of the columns a_i for $i \in I_1(I_2)$. Without loss of generality, suppose

$$A = (A_1 \; A_2). \qquad \ldots \quad (2.1)$$

Denote by $\bar{A}$ a $(q-p) \times q$ matrix whose rows form a basis of the orthocomplement of $\mathcal{M}$ in $\mathcal{R}^q$ and suppose the partitioned form of $\bar{A}$, corresponding to (2.1), is $\bar{A} = (\bar{A}_1 \; \bar{A}_2)$. Then the following can easily be recognized as an equivalent version of Theorem 1.1.

Theorem 2.1 : *A necessary and sufficient condition that the MLE $\hat{\mu}$ exist finitely is that there does not exist h such that $h \neq 0$, $h \geqslant 0$ and $\bar{A}_2 h = 0$.*

The condition stated in Theorem 2.1 may be verified in a routine manner through linear programming.

Turning to the problem of identification of I_0 and I^*, the following algorithmic steps are suggested.

Algorithmic steps :

$I_{11} \leftarrow I_1$

$k \leftarrow 1$

A $\quad A_{1k} \leftarrow$submatrix of A consisting of a_i for $i \in I_{1k}$. If rank $(A_{1k}) = p$ (which happens, in particular, when $I_{1k} = I$) go to D; otherwise go to B.

B $\quad I_{2k} \leftarrow I - I_{1k}$

$A_{2k} \leftarrow$submatrix of A consisting of a_i for $i \in I_{2k}$

$L_k \leftarrow$a matrix whose columns form a basis of the orthocomplement of column space (A_{1k}) in the p-dimensional Euclidian space.

Apply linear programming technique to maximize $1'\xi$ [where $\xi = (\ldots, \xi_i, \ldots)'$, $i \in I_{2k}$ and 1 is a vector with all elements unity] subject to $L_k' A_{2k}\xi = 0$, $\xi > 0$.

$I_{3k} \leftarrow \{i\ (\epsilon\ I_{2k}) : \xi_i > 0$ in the optimal solution$\}$. If I_{3k} is empty go to E; otherwise go to C.

C $I_{1,k+1} \leftarrow I_{1k} \bigcup I_{3k}$
$k \leftarrow k+1$
Go to A

D Conclude that the MLE $\hat{\mu}$ exists finitely.
Go to F.

E Conclude that the MLE $\hat{\mu}$ does not exist finitely, but extended MLE may obtained with $I_0 = I_{1k}$, $I^* = I_{2k}$

F End.

Since I is finite, the algorithm clearly terminates after a finite number of steps. The following example illustrates the algorithm. The proofs are given in the next section.

Example : Consider a tri-attribute situation with attributes F_1, F_2, F_3 each at two forms 1, 2. The $2^3 = 8$ form combinations (i_1, i_2, i_3) $(i_j = 1, 2;$ $j = 1, 2, 3)$ will be lexicographically ordered. Then $I = \{(1, 1, 1), (1, 1, 2),$ $(1, 2, 1), (1, 2, 2), (2, 1, 1), (2, 1, 2), (2, 2, 1), (2, 2, 2)\}$. Under the log-linear model suppose that interactions F_2F_3 and $F_1F_2F_3$ are absent. This means $V\mu = 0$, where

$$V^{(2\times 8)} = \begin{bmatrix} (1, \quad 1)\times(1, -1)\times(1, -1) \\ (1, -1)\times(1, -1)\times(1, -1) \end{bmatrix}$$

and $\times$ denotes Kronecker product. Thus $\mu \in \mathscr{M}$, the orthocomplement of row space (V) in $\mathcal{R}^8$, $p = 6$, and one may take

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \end{bmatrix},$$

the rows of A forming a basis of $\mathscr{M}$.

Let observed zero frequencies occur in the cells (1, 1, 1), (1, 2, 1), (2, 2, 1). Then $I_2 = \{(1, 1, 1), (1, 2, 1), (2, 2, 1)\}$, $I_1 = I - I_2$, $I_{11} = I_1$ and A_{11} is the 6×5 submatrix of A given by its 2nd, 4th, 5th, 6th and 8th columns. Since rank $(A_{11}) = 5\ (< p)$ one goes to step B of the algorithm to define $I_{21} = I - I_{11}$ and A_{21} as the 6×3 submatrix of A given by its 1st, 3rd and 7th columns. One may take $L_1 = (1\ 1\ 0\ 1\ 0\ 1)'$ and consider the maximization of $1'\xi$ subject to $L_1'A_{21}\xi = 0$, $\xi \geqslant 0$, where $\xi = (\xi_{111}, \xi_{121}, \xi_{221})'$. Since $L_1'A_{21} = (4, 4, 0)$, this maximization problem has an unbounded optimal solution in which $\xi_{111} = \xi_{121} = 0$ and $\xi_{221} > 0$. Hence $I_{31} = \{(2, 2, 1)\}$, $I_{12} = I_{11} \bigcup I_{31}$ and A_{12} is the 6×6 submatrix of A formed by its 2nd, 4th, 5th–8th columns. Since rank $(A_{12}) = 5\ (< p)$, one again goes to step B, defines $I_{22} = I - I_{12}$ and takes A_{22} as the 6×2 submatrix of A given by its 1st and 3rd columns. Since L_2 may be taken as $L_2 = (1\ 1\ 0\ 1\ 0\ 1)'$ and $L_2'A_{22} = (4, 4)$, the problem of maximization of $1'\xi$ subject to $L_2'A_{22}\xi = 0$, $\xi \geqslant 0$ (with $\xi = (\xi_{111}, \xi_{121})'$) yields $\xi_{111} = \xi_{121} = 0$. Hence I_{32} is empty and one may draw the consluion of step E with $I^* = \{(1, 1, 1), (1, 2, 1)\}$, $I_0 = I - I^*$.

In the above the algorithm terminates at the second stage. As another illustration, in the same tri-attribute situation suppose only interaction $F_1F_2F_3$ is absent. Now if zero cell frequencies occur in the cells (1, 1, 1) and (2, 2, 2) then it may be seen that the algorithm stops at the first stage yielding $I^* = \{(1, 1, 1), (2, 2, 2)\}$, $I_0 = I - I^*$.

Although in the above example, the algorithm performs quite satisfactorily, some problems may arise in dealing with large, sparse multi-way contingency tables. For example, storage of data may pose a problem. Also, the number of steps may become prohibitively large. It appears that further work should be done to settle these problems. Anyway, at least with tables of moderate size, it is expected that the present algorithm will be helpful.

3. Proofs

Lemma 3.1 : *For each k, there exists $y_k\ (> 0)$ such that $A_{1k}y_k = An$.*

Proof: Let $n^{(1)} = (\ldots, n_i, \ldots)'$, $i \in I_1$. Then $n^{(1)} > 0$ and trivially $A_{11}n^{(1)} = An$, i.e. the result holds for $k = 1$ with $y_1 = n^{(1)}$. To apply the method of induction, suppose the result holds for $k = m$. Then there exists $y_m\ (> 0)$ satisfying

$$A_{1m}\, y_m = An. \qquad \ldots \quad (3.1)$$

Denote by A_{3m} the submatrix of A_{2m} consisting of the columns a_i for $i \in I_{3m}$. Also if $\xi_m = (\ldots, \xi_{im}, \ldots)'$, $i \in I_{2m}$ be the optimal solution of the linear programming problem at the m-th stage, write $\xi_m^* = (\ldots, \xi_{im}, \ldots)'$, $i \in I_{3m}$. Clearly, $A_{2m}\xi_m = A_{3m}\xi_m^*$ and $\xi_m^* > 0$. Now,

$$L_m' A_{3m}\xi_m^* = L_m' A_{2m}\xi_m = 0.$$

Since the columns of L_m form a basis of the orthocomplement of column space (A_{1m}) it follows that there exists g_m such that

$$A_{3m}\xi_m^* = A_{1m}g_m. \qquad \ldots \quad (3.2)$$

As $\underset{\alpha \to 0+}{\text{Lt}} \alpha g_m = 0$, for sufficiently small positive α_0, $y_m - \alpha_0 g_m > 0$. By (3.1), (3.2),

$$A_{1m}(y_m - \alpha_0 g_m) + A_{3m}(\alpha_0 \xi_m^*) = A_{1m}y_m = An$$

i.e.
$$A_{1,m+1}y_{m+1} = An,$$

where $A_{1,m+1} = (A_{1m} \; A_{3m})$ and $y_{m+1} = \begin{pmatrix} y_m - \alpha_0 g_m \\ \alpha_0 \xi_m^* \end{pmatrix} (> 0)$.

Thus the lemma follows induction. Q.E.D.

Theorem 3.1 : *The step D in the algorithm leads to a correct decision.*

Proof : Step D is reached if $\text{rank}(A_{1k}) = p$ for some k, say for $k = m$. Clearly, by Lemma 3.1, there exists $y_m (> 0)$ such that $A_{1m}y_m = An$. Now, as $\text{rank}(A_{1m}) = p = \text{rank}(A)$, defining as usual A_{2m} as a submatrix of A consisting of the columns a_i for $i \in I - I_{1m}$, there exists a matrix Z such that $A_{2m} = A_{1m}Z$. Let $1 = (1, 1, \ldots, 1)'$, with as many components as the number of columns of A_{2m}. Since $\underset{\alpha \to 0+}{\text{Lt}} \alpha Z1 = 0$, for sufficiently small positive α_0, $y_m - \alpha_0 Z1 > 0$. Then

$$A_{1m}(y_m - \alpha_0 Z1) + A_{2m}(\alpha_0 1) = A_{1m}(y_m - \alpha_0 Z1) + A_{1m}Z(\alpha_0 1) = An,$$

i.e. $Ay = An$, where $y = \begin{pmatrix} y_m - \alpha_0 Z1 \\ \alpha_0 1 \end{pmatrix} > 0$, and the result follows by Theorem 2.2 of Haberman (1974). Q.E.D.

Theorem 3.2 : *The step E in the algorithm leads to a correct decision.*

Proof : The proof employs the following lemma from Gale (1960, p. 49) :

Lemma 3.2 : *Given any matrix T either there exists $t \geqslant 0$ ($t \neq 0$) such that $T't \leqslant 0$ or there exists $g \geqslant 0$ such that $Tg > 0$.*

Now observe that step E is reached provided for some k, I_{2k} is nonempty but I_{3k} is empty. Let this happen for $k = m$. Denote by M_m a matrix whose columns form an orthonormal basis of column space (A_{1m}) and suppose, without loss of generality, the columns of L_m are orthonormal. Then the matrix $(M_m \; L_m)$ is orthogonal and there exist matrices B_1, B_2 such that

$$A_{2m} = M_m B_1 + L_m B_2. \qquad \ldots \quad (3.3)$$

Since I_{3m} is empty, $L_m' A_{2m} \xi = 0$, $\xi \geq 0$ imply $\xi = 0$. Also by (3.3), $L_m' A_{2m} = B_2$. Therefore, an application of Lemma 3.2 shows that there exists w such that

$$B_2' w > 0. \quad \ldots$$

With $A = (A_{1m} \; A_{2m})$, let $\mu_0 = -A' L_m w$. Clearly $\mu_0 \in \mathcal{M}$. By (3.3),

$$\mu_0 = -\begin{pmatrix} A_{1m}' \\ A_{2m}' \end{pmatrix} L_m w = \begin{pmatrix} 0 \\ -B_2' w \end{pmatrix}, \quad \ldots \quad (3.5)$$

i.e. by (3.4), $\mu_0 \leqslant 0$, $\mu_0 \neq 0$. From the above it is also clear that $n'\mu_0 = 0$ since the positive elements of n correspond to zero elements of μ_0. Thus μ_0 belongs to the set C defined in Section 1. Therefore, C is nonempty and by Theorem 1.1, non-existence of (finite) MLE μ follows. Obviously, by (3.4), (3.5), $J(\mu_0) = I_{2m}$, which yields $I_{2m} \subset I^*$.

Again, for any $\mu \in C$, one can write $\mu = A'\Phi = \begin{pmatrix} A_{1m}' \Phi \\ A_{2m}' \Phi \end{pmatrix}$ for some Φ, with $A_{1m}' \Phi \leqslant 0$ as $\mu \leqslant 0$. If $y_m (> 0)$ be as in Lemma 3.1, the condition $n'\mu = 0$ now yields $y_m' A_{1m}' \Phi = 0$, whence clearly $A_{1m}' \Phi = 0$. This shows that for no $\mu \in C$, $J(\mu)$ contains any element of I_{1m}. Consequently, $I_{1m} \subset I_0$, which together with the fact $I_{2m} \subset I^*$, proves that $I_{1m} = I_0$, $I_{2m} = I^*$. Q.E.D.

Acknowledgement. The author is grateful to Professor J. K. Ghosh, Indian Statistical Institute, for proposing the problem and offering many constructive suggestions. Thanks are also due to Dr. A. B. Raha, Indian Statistical Institute, and a referee for their helpful suggestions.

References

Birch, M. M. (1963): Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc.*, Ser. B, **25**, 220-233.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975): *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, Massachusetts.

Gale, D. (1960): *The Theory of Linear Economic Models*, McGraw Hill, London.

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

——— (1978): *Analysis of Qualitative Data*, Vol. 1. Academic Press, New York.

——— (1979): *Analysis of Qualitative Data*, Vol. 2, Academic Press, New York.

Paper received : December, 1984.

Revised : November, 1986.