Obtaining the Maximum Likelihood Estimates in Incomplete $R \times C$ Contingency Tables Using a Poisson Generalized Linear Model

Stuart R. LIPSITZ, Michael PARZEN, and Geert MOLENBERGHS

This article describes estimation of the cell probabilities in an $R \times C$ contingency table with ignorable missing data. Popular methods for maximizing the incomplete data likelihood are the EM-algorithm and the Newton–Raphson algorithm. Both of these methods require some modification of existing statistical software to get the MLEs of the cell probabilities as well as the variance estimates. We make the connection between the multinomial and Poisson likelihoods to show that the MLEs can be obtained in any generalized linear models program without additional programming or iteration loops.

Key Words: EM-algorithm; Ignorable missing data; Newton–Raphson algorithm; Offset.

1. INTRODUCTION

$R \times C$ contingency tables, in which the row variable and column variable are jointly multinomial, occur often in scientific applications. The problem of estimating cell probabilities from incomplete tables—that is, when either the row or column variable is missing for some of the subjects—is very common. For example, Table 1 contains such data from the Six Cities Study, a study conducted to assess the health effects of air pollution (Ware et al. 1984). The columns of Table 1 correspond to the wheezing status (no wheeze, wheeze with cold, wheeze apart from cold) of a child at age 10. The rows represent the smoking status of the child's mother (none, medium, heavy) during that time. For some individuals the maternal smoking variable is missing, while for others the child's wheezing status is missing. One of the objectives is to estimate the probabilities of the joint distribution of maternal smoking and respiratory illness.

Two popular methods for maximizing the incomplete data likelihood (assuming ignorable nonresponse in the sense of Rubin [1976]) are the EM-algorithm (Dempster,

Stuart R. Lipsitz is Associate Professor, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115. Michael Parzen is Associate Professor, Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, IL 60637 (E-mail: fm-parzen@gsbmip.uchicago.edu). Geert Molenberghs is Associate Professor, Biostatistics, Limburgs Universitair Centrum, Universitaire Campus, B-3590 Diepenbeek, Belgium.

Table 1.   Six Cities Data: Maternal Smoking Cross-Classified by Child's Wheeze Status

Maternal Smoking	Child's wheeze status			
	No Wheeze	Wheeze with Cold	Wheeze apart from Cold	Missing
None	287	39	38	279
Moderate	18	6	4	27
Heavy	91	22	23	201
Missing	59	18	26	

Laird, and Rubin 1977) and the Newton–Raphson algorithm (Hocking and Oxspring 1971). Both of these methods require some modification of existing statistical software to get the MLEs of the cell probabilities as well as the variance estimates. In this article, we establish a connection between the multinomial and Poisson likelihoods to show that the MLEs can be obtained in any generalized linear models program, such as SAS *Proc GENMOD* (SAS Institute Inc. 1993), Stata *glm* (Stata Corporation 1995), S function *glm* (Hastie and Pregibon 1993), or GLIM (Francis, Green, and Payne1993), without any additional programming or iteration loops.

In Section 2, we formulate the missing data likelihood, and the connection between the multinomial likelihood and the Poisson likelihood. Section 3 describes the Poisson generalized linear model, including the design matrix and offset. Section 4 presents the SAS Proc GENMOD code for a (2×2) table given in Little and Rubin (1987) and the S function *glm* code for the data in Table 1.

2. INCOMPLETE $R \times C$ TABLES

Suppose that two discrete random variables, Y_{i1} and Y_{i2}, are to be observed on each of N independent subjects, where Y_{i1} can take on values $1, \ldots, R$, and Y_{i2} can take on values $1, \ldots, C$. Let the probabilities of the joint distribution of Y_{i1} and Y_{i2} be denoted by

$$p_{jk} = \text{pr}[Y_{i1} = j, Y_{i2} = k],$$

for $j = 1, \ldots, R$ and $k = 1, \ldots, C$. Because all of the probabilities must sum to 1, there are $(RC - 1)$ nonredundant multinomial cell probabilities. We let $\mathbf{p}$ denote the $(RC - 1) \times 1$ probability vector of p_{jk}'s; for simplicity, we delete p_{RC}. The marginal probabilities are $p_{j+} = \text{pr}[Y_{i1} = j]$ and $p_{+k} = \text{pr}[Y_{i2} = k]$, where '+' denote summing over the subscript it replaces. The joint probability function of Y_{i1} and Y_{i2} can be written as

$$f(y_{i1}, y_{i2}|\mathbf{p}) = \prod_{j=1}^{R} \prod_{k=1}^{C} p_{jk}^{I[Y_{i1}=j, Y_{i2}=k]},$$

where $I[\cdot]$ is an indicator function.

Now, suppose, due to an ignorable missing data mechanism (Rubin 1976), some individuals have a missing value for either Y_{i1} or Y_{i2}. When there is missing data, it is

convenient to introduce two indicator random variables R_{i1} and R_{i2}, where R_{i1} equals 1 if Y_{i1} is observed and equals 0 if Y_{i1} is missing. Similarly, R_{i2} equals 1 if Y_{i2} is observed and equals 0 if Y_{i2} is missing. It is assumed that no individuals are missing both Y_{i1} and Y_{12}—that is, no individuals have $R_{i1} = R_{i2} = 0$ (if subjects are missing on both variables, they would not be considered participants in the study). The "complete" data for subject i are $(R_{i1}, R_{i2}, Y_{i1}, Y_{i2})$, with joint distribution

$$f(y_{i1}, y_{i2}, r_{i1}, r_{i2}|\mathbf{p}, \boldsymbol{\phi}) = f(r_{i1}, r_{i2}|y_{i1}, y_{i2}, \boldsymbol{\phi})f(y_{i1}, y_{i2}|\mathbf{p}), \quad (2.1)$$

where

$$f(r_{i1}, r_{i2}|y_{i1}, y_{i2}, \boldsymbol{\phi})$$

is the "missing data mechanism" with parameter vector $\boldsymbol{\phi}$. We assume that $\boldsymbol{\phi}$ is distinct from $\mathbf{p}$.

Following the nomenclature of Rubin (1976) and Little and Rubin (1987), a hierarchy of missing data mechanisms can be distinguished. First, when $f(r_{i1}, r_{i2}|y_{i1}, y_{i2}, \boldsymbol{\phi})$ is independent of both Y_{i1} and Y_{12}, then the missing data are said to be *missing completely at random* (MCAR). When $f(r_{i1}, r_{i2}|y_{i1}, y_{i2}, \boldsymbol{\phi})$ depends on the observed data, but not on the missing values, the missing data are said to be *missing at random* (MAR). Clearly, MCAR is a special case of MAR, and often no distinction is made between these two mechanisms when they are referred to as being *ignorable*. We caution, however, that the use of the term ignorable does not imply that the individuals with missing data can simply be ignored. Rather, the term ignorable is used here to indicate that it is not necessary to specify a model for the missing data mechanism to estimate $\mathbf{p}$ in a likelihood-based analysis of the data.

Our goal is to estimate $\mathbf{p}$ using likelihood methods when the missing data are MAR. If Y_{i1} or Y_{i2} is missing, individual i will not contribute (2.1) to the likelihood of the data, but will contribute (2.1) summed over the possible values of Y_{i1} or Y_{i2}. In particular, if Y_{it} is missing ($t = 1$ or 2), individual i contributes

$$\sum_{y_{it}} f(r_{i1}, r_{i2}|y_{i1}, y_{i2}, \boldsymbol{\phi})f(y_{i1}, y_{i2}|\mathbf{p}) \quad (2.2)$$

to the likelihood. Then, the full likelihood can be written as

$$L(\boldsymbol{\phi}, \mathbf{p}) = L_1(\boldsymbol{\phi}, \mathbf{p})L_2(\boldsymbol{\phi}, \mathbf{p})L_3(\boldsymbol{\phi}, \mathbf{p}),$$

where

$$L_1(\boldsymbol{\phi}, \mathbf{p}) = \prod_{i=1}^{N} \left[f(r_{i1} = 1, r_{i2} = 1|y_{i1}, y_{i2}, \boldsymbol{\phi})f(y_{i1}, y_{i2}|\mathbf{p}) \right]^{r_{i1}r_{i2}}; \quad (2.3)$$

$$L_2(\boldsymbol{\phi}, \mathbf{p}) = \prod_{i=1}^{N} \left[\sum_{y_{i1}} f(r_{i1} = 0, r_{i2} = 1|y_{i1}, y_{i2}, \boldsymbol{\phi})f(y_{i1}, y_{i2}|\mathbf{p}) \right]^{(1-r_{i1})r_{i2}}; \quad (2.4)$$

and

$$L_3(\phi, \mathbf{p}) = \prod_{i=1}^{N} \left[\sum_{y_{i2}} f(r_{i1} = 1, r_{i2} = 0 | y_{i1}, y_{i2}, \phi) f(y_{i1}, y_{i2} | \mathbf{p}) \right]^{r_{i1}(1-r_{i2})}. \qquad (2.5)$$

Now, suppose the missing data are MAR. MAR implies that $f(r_{i1} = 0, r_{i2} = 1 | y_{i1}, y_{i2}, \phi)$ in (2.4) does not depend on y_{i1}; that is,

$$f(r_{i1} = 0, r_{i2} = 1 | y_{i1}, y_{i2}, \phi) = f(r_{i1} = 0, r_{i2} = 1 | y_{i2}, \phi).$$

Then,

$$\begin{aligned}
L_2(\phi, \mathbf{p}) &= \prod_{i=1}^{N} \left[f(r_{i1} = 0, r_{i2} = 1 | y_{i2}, \phi) \sum_{y_{i1}} f(y_{i1}, y_{i2} | \mathbf{p}) \right]^{(1-r_{i1})r_{i2}} \\
&= \prod_{i=1}^{N} \left[f(r_{i1} = 0, r_{i2} = 1 | y_{i2}, \phi) f(y_{i2} | \mathbf{p}) \right]^{(1-r_{i1})r_{i2}},
\end{aligned}$$

where

$$f(y_{i2} | \mathbf{p}) = \prod_{k=1}^{C} p_{+k}^{I[Y_{i2}=k]}$$

is the marginal distribution of Y_{i2}. Similarly, MAR implies that $f(r_{i1} = 1, r_{i2} = 0 | y_{i1}, y_{i2}, \phi)$ in (2.5) does not depend on y_{i2}; that is,

$$f(r_{i1} = 1, r_{i2} = 0 | y_{i1}, y_{i2}, \phi) = f(r_{i1} = 1, r_{i2} = 0 | y_{i1}, \phi).$$

Then,

$$\begin{aligned}
L_3(\phi, \mathbf{p}) &= \prod_{i=1}^{N} \left[f(r_{i1} = 1, r_{i2} = 0 | y_{i1}, \phi) \sum_{y_{i2}} f(y_{i1}, y_{i2} | \mathbf{p}) \right]^{r_{i1}(1-r_{i2})} \\
&= \prod_{i=1}^{N} \left[f(r_{i1} = 1, r_{i2} = 0 | y_{i1}, \phi) f(y_{i1} | \mathbf{p}) \right]^{r_{i1}(1-r_{i2})},
\end{aligned}$$

where

$$f(y_{i1} | \mathbf{p}) = \prod_{j=1}^{R} p_{j+}^{I[Y_{i1}=j]}$$

is the marginal distribution of Y_{i1}.

Then, under MAR, the likelihood $L(\phi, \mathbf{p})$ factors into two components, $L(\phi, \mathbf{p}) = L(\phi)L(\mathbf{p})$, where $L(\phi)$ is a function only of ϕ, given by

$$\begin{aligned}
L(\phi) = \prod_{i=1}^{N} &\left[f(r_{i1} = 1, r_{i2} = 1 | y_{i1}, y_{i2}, \phi) \right]^{r_{i1}r_{i2}} \\
&\times \left[f(r_{i1} = 0, r_{i2} = 1 | y_{i2}, \phi) \right]^{(1-r_{i1})r_{i2}} \left[f(r_{i1} = 1, r_{i2} = 0 | y_{i1}, \phi) \right]^{r_{i1}(1-r_{i2})}, (2.6)
\end{aligned}$$

Table 2. Notation for a (3×3) Table With Missing Data

Variable Y_{i1}	Variable Y_{i2}			
	1	2	3	Missing
1	u_{11}	u_{12}	u_{13}	w_{1+}
2	u_{21}	u_{22}	u_{23}	w_{2+}
3	u_{31}	u_{32}	u_{33}	w_{3+}
Missing	z_{+1}	z_{+2}	z_{+3}	

and $L(\mathbf{p})$ is a function only of $\mathbf{p}$, given by

$$L(\mathbf{p}) = \prod_{i=1}^{N} \left[f(y_{i1}, y_{i2}|\mathbf{p})^{r_{i1}r_{i2}} f(y_{i1}|\mathbf{p})^{r_{i1}(1-r_{i2})} f(y_{i2}|\mathbf{p})^{(1-r_{i1})r_{i2}} \right]. \tag{2.7}$$

Because the likelihood factors, the maximum likelihood estimate (MLE) of ϕ can be obtained from $L(\phi)$ and the MLE of $\mathbf{p}$ can be obtained from $L(\mathbf{p})$. We note here that an appropriate MAR missing data mechanism for $f(r_{i1}, r_{i2}|y_{i1}, y_{i2}, \phi)$ is described in detail in Chen and Fienberg (1974) and Little (1985).

We can simplify the likelihood in (2.7). We let

$$u_{jk} = \sum_{i=1}^{N} R_{i1} R_{i2} I[Y_{i1} = j, Y_{i2} = k]$$

denote the number of subjects who are observed on both Y_{i1} and Y_{i2}, with response level j on Y_{i1} and level k on Y_{i2}. Also, we let

$$w_{j+} = \sum_{i=1}^{N} R_{i1}(1 - R_{i2}) I[Y_{i1} = j]$$

denote the number of subjects with response level j on Y_{i1} who are missing Y_{i2}, and let

$$z_{+k} = \sum_{i=1}^{N} (1 - R_{i1}) R_{i2} I[Y_{i2} = k]$$

denote the number of subjects with response level k on Y_{i2} who are missing Y_{i1}. The cell counts for a (3×3) table using this notation are shown in Table 2. Using this notation, the likelihood for the cell probabilities $\mathbf{p}$ in (2.7) reduces to

$$L(\mathbf{p}) = \left[\prod_{j=1}^{R} \prod_{k=1}^{C} p_{jk}^{u_{jk}} \right] \left[\prod_{j=1}^{R} p_{j+}^{w_{j+}} \right] \left[\prod_{k=1}^{C} p_{+k}^{z_{+k}} \right]; \tag{2.8}$$

that is, the complete cases (subjects with $R_{i1} = R_{i2} = 1$) contribute

$$L_1(\mathbf{p}) = \left[\prod_{j=1}^{R} \prod_{k=1}^{C} p_{jk}^{u_{jk}} \right], \tag{2.9}$$

a multinomial likelihood with sample size 'u_{++}' and $(RC - 1)$ probabilities $\mathbf{p}$; subjects observed only on Y_{i1} contribute

$$L_2(\mathbf{p}) = \left[\prod_{j=1}^{R} p_{j+}^{w_{j+}} \right], \qquad (2.10)$$

a multinomial likelihood with sample size 'w_{++}' and $(R-1)$ nonredundant probabilities $\{p_{1+}, ... p_{R-1,+}\}$; and subjects observed only on Y_{i2} contribute

$$L_3(\mathbf{p}) = \left[\prod_{k=1}^{C} p_{+k}^{z_{+k}} \right], \qquad (2.11)$$

a multinomial likelihood with sample size 'z_{++}' and $(C - 1)$ nonredundant probabilities $\{p_{+1}, \ldots p_{+,C-1}\}$.

First, we consider the contribution to the likelihood from the complete cases (u_{jk}'s). It is a well-known fact that a multinomial likelihood can be written as a Poisson likelihood (Agresti 1990). The multinomial likelihood is written in terms of the probabilities p_{jk}, whereas the Poisson is written in terms of the expected cell counts $E(u_{jk}) = u_{++}p_{jk}$. Letting the u_{jk}'s be independent Poisson random variables, the Poisson likelihood contributed by the u_{jk}'s is

$$
\begin{aligned}
L_1(\mathbf{p}) &= e^{E(u_{++})} \prod_{j=1}^{R} \prod_{k=1}^{C} [E(u_{jk})]^{u_{jk}} \\
&= e^{(u_{++}p_{++})} \prod_{j=1}^{R} \prod_{k=1}^{C} [u_{++}p_{jk}]^{u_{jk}}, \qquad (2.12)
\end{aligned}
$$

where $p_{++} = \sum_{j=1}^{R} \sum_{k=1}^{C} p_{jk} = 1$. To constrain the Poisson likelihood to be equivalent to the multinomial likelihood with sample size u_{++}, we must restrict $E(u_{++}) = u_{++}p_{++}$ to equal u_{++}—that is, we must make sure that $p_{++} = 1$. This can be accomplished by rewriting $p_{RC} = 1 - \sum_{jk \neq RC} p_{jk}$ in the likelihood, which is most easily handled using an offset in the Poisson generalized linear model, as described in the next section.

Formally, $E(u_{jk}) = u_{++}p_{jk}$ is only true if the data are missing completely at random. However, the MLEs under missing completely at random and the weaker missing at random are identical, and it is easier to explain the connection between Poisson and multinomial likelihoods when the data are missing completely at random. Thus, we discuss obtaining the MLEs under missing completely at random, keeping in mind that the MLEs are the same as under missing at random.

Next, we consider the contribution to the likelihood from the subjects only observed on Y_{i1}—that is, the w_{j+}'s. Again, we use the connection between the multinomial likelihood and the Poisson likelihood. We again write the Poisson likelihood in terms of the expected cell counts $E(w_{j+}) = w_{++}p_{j+}$. Letting the w_{j+}'s be independent Poisson random variables, the Poisson likelihood contributed by the w_{j+}'s is

$$
\begin{aligned}
L_2(\mathbf{p}) & = e^{E(w_{++})}\prod_{j=1}^{R}[E(w_{j+})]^{w_{j+}} \\
& = e^{(w_{++}p_{++})}\prod_{j=1}^{R}[w_{++}p_{j+}]^{w_{j+}}. \quad (2.13)
\end{aligned}
$$

The Poisson likelihood will be equivalent to the multinomial if we force $E(w_{++}) = w_{++}p_{++}$ to equal w_{++}, which is again accomplished by rewriting $p_{RC} = 1 - \sum_{jk \neq RC} p_{jk}$ in the likelihood, resulting again in an offset in the part of the Poisson generalized linear model corresponding to (2.13).

Finally, we consider the contribution to the likelihood from the subjects only observed on Y_{i2}—that is, the z_{+k}'s. Letting the z_{+k}'s be independent Poisson random variables with $E(z_{+k}) = z_{++}p_{+k}$, the Poisson likelihood contributed by the z_{+k}'s is

$$
\begin{aligned}
L_3(\mathbf{p}) & = e^{E(z_{++})}\prod_{k=1}^{C}[E(z_{+k})]^{z_{+k}} \\
& = e^{(z_{++}p_{++})}\prod_{k=1}^{C}[w_{++}p_{+k}]^{z_{+k}}. \quad (2.14)
\end{aligned}
$$

Here, (2.14) will be equivalent to a multinomial likelihood when rewriting $p_{RC} = 1 - \sum_{jk \neq RC} p_{jk}$ so that $E(z_{++}) = z_{++}p_{++} = z_{++}$. As before, we need an offset in the part of the Poisson generalized linear model corresponding to (2.14).

Then, letting the u_{jk}'s, w_{j+}'s, and z_{+k}'s be independent Poisson random variables, we can maximize $L(\mathbf{p}) = L_1(\mathbf{p})L_2(\mathbf{p})L_3(\mathbf{p})$ using a Poisson generalized linear model; in particular, a Poisson linear model, as we show in the following section. To constrain $E(u_{++})$, $E(w_{++})$, and $E(z_{++})$ to equal u_{++}, w_{++} and z_{++}, respectively, we must use an offset in the generalized linear model, which will be described next.

3. THE DESIGN MATRIX FOR THE POISSON LINEAR MODEL

Suppose we stack the $(RC \times 1)$ vector $\mathbf{u} = [u_{11}, \ldots, u_{RC}]'$; the $(R \times 1)$ vector $\mathbf{w} = [w_{1+}, \ldots, w_{R+}]'$; and the $(C \times 1)$ vector $\mathbf{z} = [z_{+1}, \ldots, z_{+C}]'$ to form the outcome vector

$$
\mathbf{F} = [\mathbf{u}', \mathbf{w}', \mathbf{z}']',
$$

whose elements are independent Poisson random variables. In this section, we show that the MLE for the p_{jk}'s can be obtained from a Poisson linear model for $\mathbf{F}$ of the form

$$
E(\mathbf{F}) = X\mathbf{p} + \boldsymbol{\gamma},
$$

where $\boldsymbol{\gamma}$ is an offset, and the elements of the design matrix X are easy to obtain.

The model for the elements of $\mathbf{u}$ is

$$E(u_{jk}) = u_{++}p_{jk} \quad (jk \neq RC);$$

(3.1)

and

$$E(u_{RC}) = u_{++}\left(1 - \sum_{jk \neq RC} p_{jk}\right) = u_{++} - u_{++}\sum_{jk \neq RC} p_{jk}.$$

(3.2)

For example, if $R = C = 3$, then

$$E\begin{bmatrix} u_{11} \\ u_{12} \\ u_{13} \\ u_{21} \\ u_{22} \\ u_{23} \\ u_{31} \\ u_{32} \\ u_{33} \end{bmatrix} = u_{++} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{31} \\ p_{32} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ u_{++} \end{bmatrix}$$

(3.3)

or

$$E(\mathbf{u}) = (u_{++}A_1)\mathbf{p} + \boldsymbol{\gamma}_1,$$

where, in general, the $RC \times (RC - 1)$ matrix A_1 has its first $RC - 1$ rows equal to the $RC - 1$ identity matrix and the last row containing -1's. Also, $\boldsymbol{\gamma}_1$ is an $(RC \times 1)$ "offset" vector with all elements 0 except the last, which equals u_{++}. Finally, we write

$$E(\mathbf{u}) = X_1\mathbf{p} + \boldsymbol{\gamma}_1,$$

where $X_1 = u_{++}A_1$. The general partitioned matrix form of $E(\mathbf{u})$ is given in the appendix.

Next, consider the linear model for the subjects who have Y_{i1} observed, but are missing Y_{i2}, the w_{j+}'s. In terms of the vector $\mathbf{p}$, we have

$$E(w_{j+}) = w_{++}p_{j+} = w_{++}\sum_{k=1}^{C} p_{jk} \quad (j = 1, \ldots, R - 1);$$

and

$$\begin{aligned} E(w_{R+}) &= w_{++}\left[1 - \sum_{j=1}^{R-1} p_{j+}\right] \\ &= w_{++}\left[1 - \sum_{j=1}^{R-1}\sum_{k=1}^{C} p_{jk}\right] \\ &= w_{++} - w_{++}\left[\sum_{j=1}^{R-1}\sum_{k=1}^{C} p_{jk}\right], \end{aligned}$$

where w_{++} becomes an offset. When $R = C = 3$ as in Table 1, we have

$$
\begin{bmatrix} E(w_{1+}) \\ E(w_{2+}) \\ E(w_{3+}) \end{bmatrix} = w_{++} \begin{bmatrix} p_{1+} \\ p_{2+} \\ p_{3+} \end{bmatrix}
$$

$$
= w_{++} \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ -1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{31} \\ p_{32} \end{bmatrix}
$$

$$
+ \begin{bmatrix} 0 \\ 0 \\ w_{++} \end{bmatrix}, \tag{3.4}
$$

or $E(\mathbf{w}) = (w_{++}A_2)\mathbf{p} + \boldsymbol{\gamma}_2$. In general, the $R \times (RC - 1)$ matrix A_2 has first $(R - 1)$ rows and $(R - 1)C$ columns equal to a block diagonal matrix, with the blocks being a $(1 \times C)$ vector of 1's. The last $(C - 1)$ columns of A_2 are all zeroes. The last row of A_2 is the negative of the sum of the first $R - 1$ rows, and thus has first $(R - 1)C$ elements equal to -1, and last $(C - 1)$ elements equal to 0. Also, $\boldsymbol{\gamma}_2$ is an $(R \times 1)$ "offset" vector with all elements 0 except the last, which equals w_{++}. Finally, we write

$$
E(\mathbf{w}) = X_2\mathbf{p} + \boldsymbol{\gamma}_2,
$$

where $X_2 = w_{++}A_2$. The general partitioned matrix form of $E(\mathbf{w})$ is given in the appendix.

Finally, consider the linear model for the subjects who have Y_{i2} observed, but are missing Y_{i1}, the z_{+k}'s. In terms of the vector $\mathbf{p}$, we have

$$
E(z_{+k}) = z_{++}p_{+k} = z_{++} \sum_{j=1}^{R} p_{jk} \quad (k = 1, \ldots, C - 1);
$$

and, similar to the above derivation for w_{R+}, we have

$$
\begin{aligned}
E(z_{+C}) &= z_{++} \left[1 - \sum_{k=1}^{C-1} p_{+k} \right] \\
&= z_{++} \left[1 - \sum_{j=1}^{R} \sum_{k=1}^{C-1} p_{jk} \right] \\
&= z_{++} - z_{++} \left[\sum_{j=1}^{R} \sum_{k=1}^{C-1} p_{jk} \right],
\end{aligned}
$$

where z_{++} becomes an offset. When $R = C = 3$ as in Table 1, we have

$$\begin{bmatrix} E(z_{+1}) \\ E(z_{+2}) \\ E(z_{+3}) \end{bmatrix} = z_{++} \begin{bmatrix} p_{+1} \\ p_{+2} \\ p_{+3} \end{bmatrix}$$

$$= z_{++} \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ -1 & -1 & 0 & -1 & -1 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{31} \\ p_{32} \end{bmatrix}$$

$$+ \begin{bmatrix} 0 \\ 0 \\ z_{++} \end{bmatrix}, \tag{3.5}$$

or $E(\mathbf{z}) = (z_{++}A_3)\mathbf{p} + \boldsymbol{\gamma}_3$. In general, the first $C - 1$ rows of the $C \times (RC - 1)$ matrix A_3 horizontally concatenates R identical matrices of the form

$$[\mathbf{I}_{C-1} \ \mathbf{0}_{C-1}],$$

where $\mathbf{I}_{C-1}$ is a $C - 1$ identity matrix, and $\mathbf{0}_{C-1}$ is a $(C - 1) \times 1$ vector of 0's. After concatenating these R matrices, the last (RCth) column is deleted. The last row of A_3 is the negative of the sum of the first $(C - 1)$ rows. Also, $\boldsymbol{\gamma}_3$ is a $(C \times 1)$ "offset" vector with all elements 0 except the last, which equals z_{++}. Finally, we write

$$E(\mathbf{z}) = X_3\mathbf{p} + \boldsymbol{\gamma}_3,$$

where $X_3 = z_{++}A_3$. The general partitioned matrix form of $E(\mathbf{z})$ is given in the appendix.

Then, in any generalized linear models program, we specify a Poisson error structure, a linear link, an outcome vector of the form

$$\mathbf{F} = \begin{bmatrix} \mathbf{u} \\ \mathbf{w} \\ \mathbf{z} \end{bmatrix},$$

a design matrix

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix},$$

and, finally, an offset

$$\boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \\ \boldsymbol{\gamma}_3 \end{bmatrix}.$$

Note that the design matrix X does not contain an intercept.

Table 3. Data from Little and Rubin (1987, p. 183)

Variable Y_{i1}	Variable Y_{i2}		
	1	2	Missing
1	100	50	30
2	75	75	60
Missing	28	60	

4. EXAMPLE

In order to illustrate our method, we use the (2×2) table with missing data found in Little and Rubin (1987), which is given here in Table 3. Little and Rubin (1987, p. 183), show the MLEs are $\widehat{p}_{11} = .28$, $\widehat{p}_{12} = .17$, $\widehat{p}_{21} = .24$, and $\widehat{p}_{22} = 0.31$. Using the results of the previous section, the Poisson linear model is

$$E \begin{bmatrix} u_{11} \\ u_{12} \\ u_{21} \\ u_{22} \\ w_{1+} \\ w_{2+} \\ z_{+1} \\ z_{+2} \end{bmatrix} = \begin{bmatrix} u_{++} & 0 & 0 \\ 0 & u_{++} & 0 \\ 0 & 0 & u_{++} \\ -u_{++} & -u_{++} & -u_{++} \\ w_{++} & w_{++} & 0 \\ -w_{++} & -w_{++} & 0 \\ z_{++} & 0 & z_{++} \\ -z_{++} & 0 & -z_{++} \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{21} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ u_{++} \\ 0 \\ w_{++} \\ 0 \\ z_{++} \end{bmatrix},$$

where $u_{++} = 300$, $w_{++} = 90$, and $z_{++} = 88$. The SAS Proc Genmod commands and selected output are given in Table 4. The estimates of the p_{jk}'s in Table 4 agree with those given in Little and Rubin (1987).

Next, we use the S function *glm* to obtain the MLEs for the Six Cities data in Table 1. The models for $E(\mathbf{u})$, $E(\mathbf{w})$ and $E(\mathbf{z})$ are given in formulas (3.3), (3.4), and (3.5), respectively, with $u_{++} = 578$, $w_{++} = 507$, and $z_{++} = 103$. The commands and output are given in Table 5.

Recall that the MLEs are the same under missing at random or missing completely at random. However, when a generalized linear model program is used to estimate the p_{jk}'s, the inverse of the observed information matrix (i.e., the inverse of the negative of the Hessian matrix) should be used to estimate the variance. The inverse of the expected information is consistent only if the data are missing completely at random, whereas the inverse of the observed information is consistent under the weaker missing at random (Efron and Hinkley 1978).

5. EXTENSIONS

5.1 Multiway Tables

In this article, we have shown how any generalized linear models program can be used to obtain the MLEs of the cell probabilities in incomplete $(R \times C)$ tables. Other methods for obtaining the MLEs, such as the EM algorithm, require additional

Table 4. SAS Proc GENMOD Commands and Selected Output for Data in Table 3

```
              data one;
                input
                count     p11    p12    p21    off;
                cards;
                    100    300      0      0      0
                     50      0    300      0      0
                     75      0      0    300      0
                     75   -300   -300   -300    300
                     30     90     90      0      0
                     60    -90    -90      0     90
                     28     88      0     88      0
                     60    -88      0    -88     88
                     ;
          proc genmod data=one;
          model count = p11 p12 p21 /
              dist=poi link=id offset=off noint;
          run;
                    /* SELECTED OUTPUT */
                 Analysis Of Parameter Estimates
        Parameter  DF  Estimate  Std Err  ChiSquare  Pr>Chi
        INTERCEPT   0    0.0000   0.0000         .         .
        P11         1    0.2795   0.0223  156.9272    0.0001
        P12         1    0.1740   0.0210   68.8169    0.0001
        P21         1    0.2387   0.0227  110.9849    0.0001
        SCALE       0    1.0000   0.0000         .         .
```

programming steps. The method described here also easily extends to estimating the cell probabilities of multiway contingency tables. Although the notation gets more involved, the MLE can be calculated in any generalized linear models program with an offset and a Poisson error structure. For example, consider data from the Muscatine Coronary Risk Factor Study (Woolson and Clarke 1984), a longitudinal study to assess coronary risk factors in 4,856 school children. Children were examined in the years 1977, 1979, and 1981. The response of interest at each time is the binary variable obesity (obese, not obese). We let Y_{it} be the binary random variable (1=obese, 0=not obese) at time t, $t = 1$ for 1977, $t = 2$ for 1979, and $t = 3$ for 1981. We are interested in estimating the joint probabilities

$$p_{jk\ell} = \text{pr}[Y_{i1} = j, Y_{i2} = k, Y_{i3} = \ell],$$

$j, k, \ell = 0, 1$. If there were no missing data, we would have a $(2 \times 2 \times 2)$ contingency table. Unfortunately, 3,086 (63.6%) of the subjects are observed at some subset of the three occasions. The data are given in Table 6.

Table 5.   S function glm commands for Six Cities data in Table 1

	count	p11	p12	p13	p21	p22	p23	p31	p32	offs
1	287	528	0	0	0	0	0	0	0	0
2	39	0	528	0	0	0	0	0	0	0
3	38	0	0	528	0	0	0	0	0	0
4	18	0	0	0	528	0	0	0	0	0
5	6	0	0	0	0	528	0	0	0	0
6	4	0	0	0	0	0	528	0	0	0
7	91	0	0	0	0	0	0	528	0	0
8	22	0	0	0	0	0	0	0	528	0
9	23	-528	-528	-528	-528	-528	-528	-528	-528	528
10	279	507	507	507	0	0	0	0	0	0
11	27	0	0	0	507	507	507	0	0	0
12	201	-507	-507	-507	-507	-507	-507	0	0	507
13	59	103	0	0	103	0	0	103	0	0
14	18	0	103	0	0	103	0	0	103	0
15	26	-103	-103	0	-103	-103	0	-103	-103	103

```
Call:
glm(formula = counts   -1 + p11 + p12 + p13 + p21 + p22 + p23 + p31 + p32
+
     offset(offs), family = poisson(link = identity))


Coefficients:
      p11          p12         p13          p21          p22          p23         p31
      p32
0.4747281  0.07005741 0.0741947 0.03273146 0.01195075 0.008743729 0.206066
0.05585518

Degrees of Freedom: 15 Total; 7 Residual
Residual Deviance: 36.00067
```

Table 6. Table of data from Muscatine Coronary Risk Factor Study

Year of obesity Measurement			
1977	1979	1981	Number
0	0	0	1209
0	0	1	91
0	1	0	66
0	1	1	78
1	0	0	64
1	0	1	31
1	1	0	62
1	1	1	169
0	0	.	426
0	1	.	54
1	0	.	33
1	1	.	118
0	.	0	125
0	.	1	27
1	.	0	5
1	.	1	27
.	0	0	463
.	0	1	63
.	1	0	37
.	1	1	82
0	.	.	583
1	.	.	173
.	0	.	293
.	1	.	77
.	.	0	381
.	.	1	119

(0= not obese, 1= obese, . = missing)

Again, we assume the missing data are MAR. We let $a_{jk\ell}$ denote the number of subjects who are observed at all three times, with response level j on Y_{i1}, level k on Y_{i2}, and level ℓ on Y_{i3}; we let b_{jk+} denote the number of subjects with response level j on Y_{i1}, level k on Y_{i2}, and missing Y_{i3}; we let $c_{j+\ell}$ denote the number of subjects with response level j on Y_{i1}, level ℓ on Y_{i3} and missing Y_{i2}; we let $d_{+k\ell}$ denote the number of subjects with response level k on Y_{i2}, level ℓ on Y_{i3} and missing Y_{i1}; we let e_{j++} denote the number of subjects with response level j on Y_{i1}, and missing Y_{i2} and Y_{i3}; we let f_{+k+} denote the number of subjects with response level k on Y_{i2}, and missing Y_{i1} and Y_{i3}; and we let $g_{++\ell}$ denote the number of subjects with response level ℓ on Y_{i3}, and missing Y_{i1} and Y_{i2}.

Suppose we stack the vectors

$$\mathbf{a} = [a_{000}, a_{001}, a_{010}, a_{011}, a_{100}, a_{101}, a_{110}, a_{111}]';$$

$$\mathbf{b} = [b_{00+}, b_{01+}, b_{10+}, b_{11+}]';$$

$$\mathbf{c} = [c_{0+0}, c_{0+1}, c_{1+0}, c_{1+1}]';$$

$$\mathbf{d} = [d_{+00}, d_{+01}, d_{+10}, d_{+11}]';$$

$$\mathbf{e} = [e_{0++}, e_{1++}]';$$

$$\mathbf{f} = [f_{0++}, f_{1++}]';$$

and

$$\mathbf{g} = [g_{0++}, g_{1++}]',$$

to form the outcome vector

$$\mathbf{F} = [\mathbf{a}', \mathbf{b}', \mathbf{c}', \mathbf{d}', \mathbf{e}', \mathbf{f}', \mathbf{g}']',$$

whose elements are independent Poisson random variables. Using similar ideas as in the previous sections, the MLE for the $p_{jk\ell}$'s can be obtained from a Poisson linear model for $\mathbf{F}$ of the form

$$E(\mathbf{F}) = X\mathbf{p} + \boldsymbol{\gamma},$$

where $\boldsymbol{\gamma}$ is an offset, X is the design matrix, and $\mathbf{p} = [p_{000}, p_{001}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}]$. We delete $p_{111} = (1 - p_{000} - p_{001} - p_{010} - p_{011} - p_{100} - p_{101} - p_{110})$ from $\mathbf{p}$ since it is redundant. In particular, the model for $\mathbf{a}$ is

$$E\begin{bmatrix} a_{000} \\ a_{001} \\ a_{010} \\ a_{011} \\ a_{100} \\ a_{101} \\ a_{110} \\ a_{111} \end{bmatrix} = a_{+++} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} p_{000} \\ p_{001} \\ p_{010} \\ p_{011} \\ p_{100} \\ p_{101} \\ p_{110} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ a_{+++} \end{bmatrix}; \tag{5.1}$$

the model for $\mathbf{b}$ is

$$\begin{bmatrix} E(b_{00+}) \\ E(b_{01+}) \\ E(b_{10+}) \\ E(b_{11+}) \end{bmatrix} = b_{+++} \begin{bmatrix} p_{00+} \\ p_{01+} \\ p_{10+} \\ p_{11+} \end{bmatrix}$$

$$= b_{+++} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ -1 & -1 & -1 & -1 & -1 & -1 & 0 \end{bmatrix} \begin{bmatrix} p_{000} \\ p_{001} \\ p_{010} \\ p_{011} \\ p_{100} \\ p_{101} \\ p_{110} \end{bmatrix}$$

$$+ \begin{bmatrix} 0 \\ 0 \\ 0 \\ b_{+++} \end{bmatrix}; \tag{5.2}$$

the model for **c** is

$$
\begin{bmatrix}
E(c_{0+0}) \\
E(c_{0+1}) \\
E(c_{1+0}) \\
E(c_{1+1})
\end{bmatrix}
= c_{+++}
\begin{bmatrix}
p_{0+0} \\
p_{0+1} \\
p_{1+0} \\
p_{1+1}
\end{bmatrix}
$$

$$
= c_{+++}
\begin{bmatrix}
1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 \\
-1 & -1 & -1 & -1 & -1 & 0 & -1
\end{bmatrix}
\begin{bmatrix}
p_{000} \\
p_{001} \\
p_{010} \\
p_{011} \\
p_{100} \\
p_{101} \\
p_{110}
\end{bmatrix}
$$

$$
+ \begin{bmatrix}
0 \\
0 \\
0 \\
c_{+++}
\end{bmatrix} ;
\tag{5.3}
$$

the model for **d** is

$$
\begin{bmatrix}
E(d_{+00}) \\
E(d_{+01}) \\
E(d_{+10}) \\
E(d_{+11})
\end{bmatrix}
= d_{+++}
\begin{bmatrix}
p_{+00} \\
p_{+01} \\
p_{+10} \\
p_{+11}
\end{bmatrix}
$$

$$
= d_{+++}
\begin{bmatrix}
1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 \\
-1 & -1 & -1 & 0 & -1 & -1 & -1
\end{bmatrix}
\begin{bmatrix}
p_{000} \\
p_{001} \\
p_{010} \\
p_{011} \\
p_{100} \\
p_{101} \\
p_{110}
\end{bmatrix}
$$

$$
+ \begin{bmatrix}
0 \\
0 \\
0 \\
d_{+++}
\end{bmatrix} ;
\tag{5.4}
$$

the model for **e** is

$$\left[\begin{array}{c} E(e_{0++}) \\ E(e_{1++}) \end{array}\right] = e_{+++} \left[\begin{array}{c} p_{0++} \\ p_{1++} \end{array}\right]$$

$$= e_{+++} \left[\begin{array}{ccccccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ -1 & -1 & -1 & -1 & 0 & 0 & 0 \end{array}\right] \left[\begin{array}{c} p_{000} \\ p_{001} \\ p_{010} \\ p_{011} \\ p_{100} \\ p_{101} \\ p_{110} \end{array}\right]$$

$$+ \left[\begin{array}{c} 0 \\ e_{+++} \end{array}\right]; \hspace{3cm} (5.5)$$

the model for **f** is

$$\left[\begin{array}{c} E(f_{+0+}) \\ E(f_{+1+}) \end{array}\right] = f_{+++} \left[\begin{array}{c} p_{+0+} \\ p_{+1+} \end{array}\right]$$

$$= f_{+++} \left[\begin{array}{ccccccc} 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ -1 & -1 & 0 & 0 & -1 & -1 & 0 \end{array}\right] \left[\begin{array}{c} p_{000} \\ p_{001} \\ p_{010} \\ p_{011} \\ p_{100} \\ p_{101} \\ p_{110} \end{array}\right] + \left[\begin{array}{c} 0 \\ f_{+++} \end{array}\right];$$

$$(5.6)$$

and the model for **g** is

$$\left[\begin{array}{c} E(g_{++0}) \\ E(g_{++1}) \end{array}\right] = g_{+++} \left[\begin{array}{c} p_{++0} \\ p_{++1} \end{array}\right]$$

$$= g_{+++} \left[\begin{array}{ccccccc} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ -1 & 0 & -1 & 0 & -1 & 0 & -1 \end{array}\right] \left[\begin{array}{c} p_{000} \\ p_{001} \\ p_{010} \\ p_{011} \\ p_{100} \\ p_{101} \\ p_{110} \end{array}\right] + \left[\begin{array}{c} 0 \\ g_{+++} \end{array}\right].$$

$$(5.7)$$

The SAS Proc Genmod commands for obtaining **p** and selected output are given in Table 7.

5.2   Other Models for p

The method can be used to obtain the MLE for any linear model for **p**, in any generalized linear model program, without any additional programming. For example,

Table 7.   SAS Proc GENMOD Commands and Selected Output for Data in Table 6

```
data one;
input
   count p000 p001 p010 p011 p100 p101 p110 off margtot;
p000 = p000*margtot; p001 = p001*margtot;
p010 = p010*margtot; p011 = p011*margtot;
p100 = p100*margtot; p101 = p101*margtot;
p110 = p110*margtot; p111 = p111*margtot;
off = off*margtot;
cards;
      1209    1    0    0    0    0    0    0   0   1770
        91    0    1    0    0    0    0    0   0   1770
        66    0    0    1    0    0    0    0   0   1770
        78    0    0    0    1    0    0    0   0   1770
        64    0    0    0    0    1    0    0   0   1770
        31    0    0    0    0    0    1    0   0   1770
        62    0    0    0    0    0    0    1   0   1770
       169   -1   -1   -1   -1   -1   -1   -1   1   1770
       426    1    1    0    0    0    0    0   0    631
        54    0    0    1    1    0    0    0   0    631
        33    0    0    0    0    1    1    0   0    631
       118   -1   -1   -1   -1   -1   -1    0   1    631
       125    1    0    1    0    0    0    0   0    184
        27    0    1    0    1    0    0    0   0    184
         5    0    0    0    0    1    0    1   0    184
        27   -1   -1   -1   -1   -1    0   -1   1    184
       463    1    0    0    0    1    0    0   0    645
        63    0    1    0    0    0    1    0   0    645
        37    0    0    1    0    0    0    1   0    645
        82   -1   -1   -1    0   -1   -1   -1   1    645
       583    1    1    1    1    0    0    0   0    756
       173   -1   -1   -1   -1    0    0    0   1    756
       293    1    1    0    0    1    1    0   0    370
        77   -1   -1    0    0   -1   -1    0   1    370
       381    1    0    1    0    1    0    1   0    500
       119   -1    0   -1    0   -1    0   -1   1    500
         ;

   proc genmod data=one;
   model count = p000 p001 p010 p011 p100 p101 p110 /
          dist=poi link=id offset=off noint;
   run;
```

Table 7.   continued

```
/* SELECTED OUTPUT */


Analysis Of Parameter Estimates


Parameter  DF  Estimate  Std Err  ChiSquare  Pr¿Chi


INTERCEPT  0    0.0000    0.0000        .          .
P000       1    0.6633    0.0078   7190.7713   0.0001
P001       1    0.0578    0.0048    143.2562   0.0001
P010       1    0.0348    0.0037     86.5835   0.0001
P011       1    0.0439    0.0042    110.6773   0.0001
P100       1    0.0356    0.0039     83.4538   0.0001
P101       1    0.0207    0.0033     40.2450   0.0001
P110       1    0.0357    0.0039     82.6427   0.0001
SCALE      0    1.0000    0.0000        .          .
```

the marginal homogeneity model (Firth 1989) is a linear model for $\mathbf{p}$, say $\mathbf{p} = D\boldsymbol{\beta}$. Then $\boldsymbol{\beta}$ can be calculated using any generalized linear model program without additional programming since the model for $\mathbf{F}$ is still linear

$$
\begin{aligned}
E(\mathbf{F}) &= X\mathbf{p} + \boldsymbol{\gamma} \\
&= X(D\boldsymbol{\beta}) + \boldsymbol{\gamma} \\
&= X^*\boldsymbol{\beta} + \boldsymbol{\gamma},
\end{aligned}
$$

where $X^* = XD$.

With additional programming in a generalized linear model program, our method can be used to fit a nonlinear model for $\mathbf{p}$. For example, log-linear models of the form $\mathbf{p} = \exp(D\boldsymbol{\beta})$, give rise to

$$
\begin{aligned}
E(\mathbf{F}) &= X\mathbf{p} + \boldsymbol{\gamma} \\
&= X[\exp(D\boldsymbol{\beta})] + \boldsymbol{\gamma}.
\end{aligned}
\tag{5.8}
$$

Unfortunately, without additional programming, the generalized linear models programs discussed in the introduction only allow Poisson regression models of the form

$$
E(\mathbf{F}) = g(A\boldsymbol{\beta} + \boldsymbol{\gamma}),
$$

for a design matrix A, and where $g(\cdot)$ is the linear, exponential, or inverse functions. Unfortunately, (5.8) does not fall in this class, and will require additional programming. To use a generalized linear models program, one must write a macro that calculates (5.8) as well as the derivative of (5.8) with respect to $\boldsymbol{\beta}$. Alternatively, a maximization procedure based on ideas similar to this article is discussed in Molenberghs and Goetghebeur (1997).

APPENDIX

In this appendix, we give the partitioned matrix form for the models for $E(\mathbf{u})$, $E(\mathbf{w})$, and $E(\mathbf{z})$. We define the following notation: $\mathbf{I}_a$ is an $a \times a$ identity matrix, $\mathbf{1}_a$ is an $a \times 1$ vector of 1's, $\mathbf{0}_{a,b}$ is an $a \times b$ matrix of 0's, and $\otimes$ is the "direct product."

In partitioned matrix form, the models are

$$E(\mathbf{u}) = u_{++} \left[\frac{\mathbf{I}_{RC-1}}{-\mathbf{1}'_{RC-1}} \right] \mathbf{p} + \left[\frac{\mathbf{0}_{RC-1,1}}{u_{++}} \right];$$

$$E(\mathbf{w}) = w_{++} \left[\begin{array}{c|c} \mathbf{I}_{R-1} \otimes \mathbf{1}'_C & \mathbf{0}_{R-1,C-1} \\ \hline -\mathbf{1}'_{(R-1)C} & \mathbf{0}_{1,C-1} \end{array} \right] \mathbf{p} + \left[\frac{\mathbf{0}_{R-1,1}}{w_{++}} \right];$$

and

$$E(\mathbf{z}) = z_{++} \left\{ \left[\mathbf{1}'_R \otimes \left(\begin{array}{c|c} \mathbf{I}_{C-1} & \mathbf{0}_{C-1,1} \\ \hline -\mathbf{1}'_{C-1} & 0 \end{array} \right) \right] \left(\frac{\mathbf{I}_{RC-1}}{\mathbf{0}_{1,RC-1}} \right) \right\} \mathbf{p} + \left[\frac{\mathbf{0}_{C-1,1}}{z_{++}} \right].$$

ACKNOWLEDGMENTS

[Received March 1996. Revised November 1997.]

REFERENCES

Agresti, A. (1990), *Categorical Data Analysis*, New York: Wiley.

Chen, T., and Fienberg, S. E.(1974), "Two-Dimensional Contingency Tables With Both Completely and Partially Cross-Classified Data," *Biometrics*, 30, 629–642.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 39, 1–38.

Efron, B., and Hinkley, D. (1978), "Assessing the Accuracy of the Maximum Likelihood Estimator: Observed Versus Expected Fisher Information," *Biometrika*, 65, 457–481.

Firth, D. (1989), "Marginal Homogeneity and the Superposition of Latin Squares," *Biometrika*, 76, 179–182.

Francis, B., Green, M., and Payne, C. (1993), *The GLIM system. Release 4 Manual*, New York: Clarendon Press.

Hastie, T.J., and Pregibon, D. (1993), "Generalized Linear Models," in *Statistical Models in S*, eds. J.M. Chambers and T.J. Hastie, Pacific Grove, CA: Wadsworth & Brooks.

Hocking, R.R., and Oxspring, H. H. (1971), "Maximum Likelihood Estimation With Incomplete Multinomial Data," *Journal of the American Statistical Association*, 63, 65–70.

Little, R.J.A. (1985), "Nonresponse Adjustments in Longitudinal Surveys: Models for Categorical Data," in *Proceedings of the 45th Session of the International Statistics Institute*, Book 3, 15.1-1–15.1-17.

Little, R.J.A., and Rubin, D.B. (1987), *Statistical Analysis with Missing Data*, New York: Wiley.

Molenberghs, G., and Goetghebeur, E. (1997), "Simple Fitting Algorithms for Incomplete Categorical Data," *Journal of the Royal Statistical Society*, Ser. B, 59, 401–414.

Rubin, D.B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581–592.

SAS Institute, Inc.(1993), "SAS Technical Report P-243," *SAS/STAT Software: The GENMOD Procedure*, (Release 6.09), Cary, NC: SAS Institute Inc.

Stata Corporation (1995), *Stata Reference Manual*, (Release 4.0, vol. 2), College Station, TX, pp. 399–410.

Ware, J.H., Dockery, D.W., Spiro, A. III, Speizer, F.E., and Ferris, B.G., Jr. (1984), "Passive Smoking, Gas Cooking, and Respiratory Health of Children Living in Six Cities," *American Review of Respiratory Diseases*, 129, 366–374.

Woolson, R.F., and Clarke, W.R. (1984), "Analysis of Categorical Incomplete Longitudinal Data," *Journal of the Royal Statistical Society*, Ser. A, 147, 87–99.