

11-1-2008

Analyzing Incomplete Categorical Data: Revisiting Maximum Likelihood Estimation (Mle) Procedure

Hoo Ling Ping

The University of Nottingham, lphoo_04@yahoo.com

M. Ataharul Islam

University of Dhaka, Bangladesh, mataharul@yahoo.com

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Ping, Hoo Ling and Islam, M. Ataharul (2008) "Analyzing Incomplete Categorical Data: Revisiting Maximum Likelihood Estimation (Mle) Procedure," *Journal of Modern Applied Statistical Methods*: Vol. 7 : Iss. 2 , Article 14.
DOI: 10.22237/jmasm/1225512780

Analyzing Incomplete Categorical Data: Revisiting Maximum Likelihood Estimation (MLE) Procedure

Hoo Ling Ping
The University of Nottingham,
Malaysia

M. Ataharul Islam
University of Dhaka,
Bangladesh

Incomplete data poses formidable difficulties in the application of statistical techniques and requires special procedures to handle. The most common ways to solve this problem are by ignoring, truncating, censoring or collapsing those data, but these may lead to inappropriate conclusions because those data might contain important information. Most of the research for estimating cell probabilities involving incomplete categorical data is based on the EM algorithm. A likelihood approach is employed for estimating cell probabilities for missing values and makes comparisons between maximum likelihood estimation (MLE) and the EM algorithm. The MLE can provide almost the same estimates as that of the EM algorithm without any loss of properties. Results are compared for different distributional assumptions. Using clinical trial results from a group of 59 epileptics, results from the application of MLE and EM algorithm are compared and the advantages of MLE are highlighted.

Key words: Incomplete categorical data, maximum likelihood estimation (MLA), EM algorithm, multinomial distribution, binomial distribution, Poisson distribution, Newton-Raphson method.

Introduction

Incomplete data is referred to as data in which entries are missing, were a prior zero or are undetermined (Fienberg, 1980). Incomplete data is one of the main obstacles to researchers; this is especially true in the case of incomplete categorical data. The most common ways to solve this problem are by ignoring, truncating, censoring or collapsing those data; however, such procedures may lead to confusion and/or inappropriate conclusions because those data might contain important information.

Little & Rubin (1987) defined the missing data mechanisms as *ignorable missing data mechanism* and *non-ignorable missing data mechanism*. The ignorable missing data mechanism involves process missing completely

at random (MCAR) and missing at random (MAR). When the missingness is independent of both unobserved and observed data, the non-response process is named as MCAR. However, if the missingness is independent of the unobserved measurement conditionally on the observed data, the non-response process is called MAR. Non-ignorable missing data mechanisms involve *informative* process. When the process is neither MCAR nor MAR, then the process is termed informative. This article considers the missing data mechanism as a non-ignorable missing data mechanism.

The problem of estimation for incomplete contingency tables under the quasi-independence model was examined by Fienberg (1970), who used the maximum likelihood estimation (MLE) procedure. Similarly, MLE for the Poisson and Multinomial sampling distributions for the incomplete contingency tables in the presence of missing row and missing column data were considered by Chen & Fienberg (1974). Chen & Fienberg (1976) extended their works which focused on cross-classifications containing some totally mixed up cell frequencies with multinomial sampling. In the following year, Dempster, Laird & Rubin

Hoo Ling Ping is Assistant Professor in the Department of Applied Mathematics, Faculty of Engineering. Email: lphoo_04@yahoo.com. M. Ataharul Islam is Professor in the Department of Statistics, Biostatistics and Informatics. Email: mataharul@yahoo.com.

presented MLE of incomplete data and named the algorithm EM since each iteration of the algorithm involved expectation (E) and maximization (M) steps. This method has been used extensively by other researchers especially for incomplete categorical data. Among others, Fuchs (1982), Nordheim (1984), Fay (1986), Baker & Laird (1988), and Philips (1993) have used the EM algorithm for analyzing incomplete categorical data. The EM algorithm was used to improve the convergence of the EM by incorporating the Newton-Raphson approach by Baker (1994) and Galecki & Molenberghs (2001). The EM algorithm is well developed (Lauritzen, 1995) to exploit the computational scheme of Lauritzen & Spiegelhalter (1988) to perform the E-step of EM algorithm to find MLE in hierarchical log-linear models and recursive models for contingency tables with missing data. Molenberghs & Goetchebeur (1997) presented a simple expression of the observed data log-likelihood for the EM algorithm. Tang, et. al. (2007) also found that the EM algorithm is the most widely used approach for finding the maximum likelihood estimate for incomplete data situations, but it lacks the direct provision of a measure of precision for the estimators and has a slow rate of convergence.

Because the EM algorithm was introduced, the MLE procedure was largely ignored by researchers until 1985. Stasny (1985) used MLE to process the model based on data from a Current Population Survey, and also used a Labor Force Survey to estimate gross flow data. Most recently, Lyles & Allen (2003) proposed MLE with multinomial likelihood, properly accounting for missing data and assuming that the probability of missing exposure depends on true exposure.

In this article, not only is the missing row or missing column data redistributed, but also both row and column missing data for multinomial sampling by extending the works of Chen & Fienberg (1974). Both row and column missing data are also investigated for the EM algorithm which has not been studied before. The MLE method for Poisson and Binomial sampling distributions was also examined as an extension of the works of Chen & Fienberg (1974). The binomial distribution can be

considered a special case of the Multinomial distribution. The same sampling patterns for the EM algorithm are considered here. The Newton-Raphson method was adopted in the MLE procedure to make convergence faster. Results of the MLE and EM algorithm are compared and the advantages of MLE are highlighted.

This article is organized as follows: data taken from Diggle, Liang & Zeger (1994) is described, followed by the formulation of the MLE and EM algorithms. Finally, results are discussed, testing independence is presented and conclusions are put forth.

Methodology

The data considered herein was referred from Diggle, Liang & Zeger (1994) based on a clinical trial of 59 epileptics. For each patient, the number of epileptic seizures was recorded during an eight week baseline period. Patients were then randomized to either a treatment group with anti-epileptic drug progabide (0) or to a placebo group (1) and the number of seizures was recorded in four consecutive two-week intervals. Table 1 shows the 2x2 artificial incomplete contingency table; rows refer to the treatment and columns refer to the results of the last treatment for the patient. The result of treatment is recorded as Y.

Maximum likelihood estimation (MLE), Poisson and multinomial distribution

Chen & Fienberg (1974) considered the MLE for incomplete contingency tables when missing row and column data existed. Their works are extended by considering incomplete contingency tables where either row or column, or both row and column are missing.

Let the fully cross-classified count for the $(i, j)^{\text{th}}$ cell of an $r \times c$ contingency table be x_{ij} , R_i ($i = 1, 2, \dots, r$) is the count of the partially classified individuals corresponding to the i^{th} row, C_j ($j = 1, 2, \dots, c$) is the count of the partially classified individuals corresponding to the j^{th} column, and D is the count of missing in both row and column. (See Figure 1.) Therefore the total sample size is:

$$N = \sum_{ij} x_{ij} + \sum_i R_i + \sum_j C_j + D = x_{++} + R_+ + C_+ + D. \quad (1)$$

ANALYZING INCOMPLETE CATEGORICAL DATA: REVISITING MLE PROCEDURE

Table1: Incomplete data

a) No missing on treatment and Y

		Y		Total
		≤ 5	> 5	
Treatment	0	13	7	20
	1	12	7	19
Total		25	14	39

b) Missing Y, treatment, Y and treatment

		Missing Y	Y		Total
Treatment	0	3	Missing	Yes	4
	1	7	Treatment	No	
Total		10	2	2	

	Missing Y	Total
Missing Treatment	6	
Total		6

Figure 1: Illustration for complete observed and incomplete data

a) Complete observed data

	Row				Total
Column	x_{11}	x_{12}	...	x_{1c}	x_{1+}
	x_{21}	x_{22}	...	x_{2c}	x_{2+}
	\vdots	\vdots	\ddots	\vdots	\vdots
	x_{r1}	x_{r2}	...	x_{rc}	x_{c+}
Total	x_{+1}	x_{+2}		x_{+c}	x

b) Incomplete units

	Missing column	Column				Total	
Row	R_{1+}	Missing row	C_{+1}	C_{+2}	...	C_{+c}	C
	R_{2+}						
	\vdots						
	R_{r+}						
Total	R						

	Missing column	Total
Missing row	D	
Total		D

When the original sampling scheme is Poisson with an expected value m_{ij} for the $(i, j)^{th}$ cell, parameters associated with the cells (illustrated in Table 2) where $\lambda_{1(i)}$, $\lambda_{2(j)}$ and λ are referred to the probabilities of losing its row, column, and both row and column identity, respectively. The cell probability of multinomial sampling for a completely classified $(i, j)^{th}$ cell is π_{ij} and $\sum_i \sum_j \pi_{ij} = 1$. By replacing π_{ij} with m_{ij} , the likelihood function for Poisson is

$$\exp\left(-\sum_i \sum_j m_{ij}\right) \prod_i \prod_j \left[(1 - \lambda_{1(i)} - \lambda_{2(j)} - \lambda) m_{ij} \right]^{x_{ij}}$$

$$\prod_i (\lambda_{1(i)} m_{i+})^{R_i} \prod_j (\lambda_{2(j)} m_{+j})^{C_j} (\lambda m_{++})^D \tag{2}$$

where

$$m_{i+} = \sum_j m_{ij}, m_{+j} = \sum_i m_{ij}, \text{ and } m_{++} = \sum_i \sum_j m_{ij}$$

for all i and j .

Equation (2) is a product of functions f_1 and f_2 defined as follows

$$f_1 = \left[\prod_i \prod_j (1 - \lambda_{1(i)} - \lambda_{2(j)} - \lambda)^{x_{ij}} \right] \tag{3}$$

$$\left[\prod_i \lambda_{1(i)}^{R_i} \right] \left[\prod_j \lambda_{2(j)}^{C_j} \right] \left[\lambda^D \right]$$

and

$$f_2 = \frac{\exp\left(-\sum_i \sum_j m_{ij}\right) \left[\prod_i \prod_j m_{ij}^{x_{ij}} \right]}{\left[\prod_i m_{i+}^{R_i} \right] \left[\prod_j m_{+j}^{C_j} \right] \left[m_{++}^D \right]} \tag{4}$$

when considering the unrestricted log linear model where

$$\log m_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \tag{5}$$

Therefore,

$$\log f_2 = -\sum_i \sum_j m_{ij} + \sum_i \sum_j x_{ij} \log m_{ij} + \sum_i R_i \log m_{i+} + \sum_j C_j \log m_{+j} + D \log m_{++}$$

$$= -\sum_i \sum_j e^{\mu + \alpha_i + \beta_j + \gamma_{ij}} + \sum_i \sum_j x_{ij} (\mu + \alpha_i + \beta_j + \gamma_{ij}) + \sum_i R_i (\mu + \alpha_i + \beta_+ + \gamma_{i+})$$

$$+ \sum_j C_j (\mu + \alpha_+ + \beta_j + \gamma_{j+}) + D (\mu + \alpha_+ + \beta_+ + \gamma_{++}) \tag{6}$$

Differentiating (6) with μ , α_i , β_j and γ_{ij} , results in:

$$\frac{\partial \log f_2}{\partial \mu} = \frac{-\sum_i \sum_j e^{\mu + \alpha_i + \beta_j + \gamma_{ij}}}{+\sum_i \sum_j x_{ij} + \sum_i R_i + \sum_j C_j + D}$$

$$= -\sum_i \sum_j m_{ij} + \sum_i \sum_j x_{ij} + \sum_i R_i + \sum_j C_j + D = -m_{++} + x_{++} + R_+ + C_+ + D$$

$$\frac{\partial \log f_2}{\partial \alpha_i} = -m_{i+} + x_{i+} + R_i + \sum_j C_j \left(\frac{m_{ij}}{m_{+j}} \right) + D \left(\frac{m_{ij}}{m_{++}} \right)$$

$$\frac{\partial \log f_2}{\partial \beta_j} = -m_{+j} + x_{+j} + \sum_i R_i \left(\frac{m_{ij}}{m_{i+}} \right) + C_j + D \left(\frac{m_{ij}}{m_{++}} \right)$$

$$\frac{\partial \log f_2}{\partial \gamma_{ij}} = m_{ij} + x_{ij} + R_i \left(\frac{m_{ij}}{m_{i+}} \right) + C_j \left(\frac{m_{ij}}{m_{+j}} \right) + D \left(\frac{m_{ij}}{m_{++}} \right) \tag{7}$$

When (7) is equal to 0,

$$\hat{m}_{ij} = x_{ij} + R_i \left(\frac{\hat{m}_{ij}}{m_{i+}} \right) + C_j \left(\frac{\hat{m}_{ij}}{m_{+j}} \right) + D \left(\frac{\hat{m}_{ij}}{m_{++}} \right) \tag{8}$$

As from Chen & Fienberg (1974), (8) is not able to be solved in closed form; initial estimates of the $\{\hat{m}_{ij}\}$ will be considered as

$$m_{ij}^{(0)} = \left(\frac{x_{ij}}{x_{++}} \right) N \tag{9}$$

On the first iteration, from (8)

$$\hat{m}_{ij}^{(1)} = x_{ij} + R_i \left(\frac{\hat{m}_{ij}^{(0)}}{m_{i+}^{(0)}} \right) + C_j \left(\frac{\hat{m}_{ij}^{(0)}}{m_{+j}^{(0)}} \right) + D \left(\frac{\hat{m}_{ij}^{(0)}}{m_{++}^{(0)}} \right) \tag{10}$$

therefore on $(k+1)^{th}$ iteration,

$$\hat{m}_{ij}^{(k+1)} = x_{ij} + R_i \left(\frac{\hat{m}_{ij}^{(k)}}{m_{i+}^{(k)}} \right) + C_j \left(\frac{\hat{m}_{ij}^{(k)}}{m_{+j}^{(k)}} \right) + D \left(\frac{\hat{m}_{ij}^{(k)}}{m_{++}^{(k)}} \right) \tag{11}$$

When $k \rightarrow \infty$, $|\hat{m}_{ij}^{(k+1)} - \hat{m}_{ij}^{(k)}| \leq \epsilon$.

Table 2: Underlying probabilities for a 2x2 table

	<u>Fully Classified Table</u>		Row Supplemental Margin
	$(1 - \lambda_{1(1)} - \lambda_{2(1)} - \lambda)\pi_{11}$	$(1 - \lambda_{1(1)} - \lambda_{2(2)} - \lambda)\pi_{12}$	$\lambda_{1(1)}\pi_{1+}$
	$(1 - \lambda_{1(2)} - \lambda_{2(1)} - \lambda)\pi_{21}$	$(1 - \lambda_{1(2)} - \lambda_{2(2)} - \lambda)\pi_{22}$	$\lambda_{1(2)}\pi_{2+}$
			Missing row and column
Column Supplemental Margin	$\lambda_{2(1)}\pi_{+1}$	$\lambda_{2(2)}\pi_{+2}$	$\lambda\pi_{++}$

Since $\pi_{ij} = \frac{x_{ij}}{x_{++}}$ for complete classified multinomial sampling, from (9) results,

$$\hat{\pi}_{ij}^{(0)} = \frac{\hat{m}_{ij}^{(0)}}{N},$$

and on the (k+1)th iteration, $\hat{\pi}_{ij}^{(k+1)} = \frac{\hat{m}_{ij}^{(k+1)}}{N}$.

Poisson and binomial distributions

Now consider the complete contingency tables where there exist missing column. The fully cross-classified count for the (i, j) cell of an r x 2 contingency table is x_{ij} , and R_i ($i = 1, 2, \dots, r$) is the count of the partially classified individuals corresponding to the ith row. Therefore the total sample size is

$$N = \sum_{ij} x_{ij} + \sum_i R_i = x_{++} + R_+ \quad (12)$$

When the original sampling scheme is Poisson with expected value m_{ij} for the (i, j) cell, then the parameters associated with the cells (illustrated in Table 3) where $\lambda_{1(i)}$, is referred to the probabilities of losing its row.

The cell probability of Binomial sampling for complete classified of (i, j) cell is

π_{ij} and $\sum_i \sum_j \pi_{ij} = 1$. By replacing π_{ij} with m_{ij} , the likelihood function for Poisson is

$$\exp\left(-\sum_i \sum_j m_{ij}\right) \prod_i \prod_j \left[(1 - \lambda_{1(i)}) m_{ij}\right]^{x_{ij}} \prod_i (\lambda_{1(i)} m_{i+})^{R_i}, \quad (13)$$

where $m_{i+} = \sum_j m_{ij}$ for all i and j.

Equation (13) is a product of a function

$$f_1 = \left[\prod_i \prod_j (1 - \lambda_{1(i)})^{x_{ij}} \right] \left[\prod_i \lambda_{1(i)}^{R_i} \right] \quad (14)$$

and

$$f_2 = \exp\left(-\sum_i \sum_j m_{ij}\right) \left[\prod_i \prod_j m_{ij}^{x_{ij}} \right] \left[\prod_i m_{i+}^{R_i} \right]. \quad (15)$$

When considering the unrestricted log linear model where

$$\log m_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \quad (16)$$

then,

$$\begin{aligned} \log f_2 &= -\sum_i \sum_j m_{ij} + \sum_i \sum_j x_{ij} \log m_{ij} \\ &+ \sum_i R_i \log m_{i+} \\ &- \sum_i \sum_j e^{\mu+\alpha_i+\beta_j+\gamma_{ij}} + \sum_i \sum_j x_{ij} \\ &= (\mu + \alpha_i + \beta_j + \gamma_{ij}) \\ &+ \sum_i R_i (\mu + \alpha_i + \beta_+ + \gamma_{i+}) \end{aligned} \tag{17}$$

Differentiating (17) with μ , α_i , β_j and γ_{ij} , results in

$$\begin{aligned} \frac{\partial \log f_2}{\partial \mu} &= -\sum_i \sum_j e^{\mu+\alpha_i+\beta_j+\gamma_{ij}} + \sum_i \sum_j x_{ij} + \sum_i R_i \\ &= -\sum_i \sum_j m_{ij} + \sum_i \sum_j x_{ij} + \sum_i R_i \\ &= -m_{++} + x_{++} + R_+ \\ \frac{\partial \log f_2}{\partial \alpha_i} &= -m_{i+} + x_{i+} + R_i \\ \frac{\partial \log f_2}{\partial \beta_j} &= -m_{+j} + x_{+j} + \sum_i R_i \left(\frac{m_{ij}}{m_{+j}} \right) \\ \frac{\partial \log f_2}{\partial \gamma_{ij}} &= -m_{ij} + x_{ij} + R_i \left(\frac{m_{ij}}{m_{i+}} \right) \end{aligned} \tag{18}$$

and, when (18) is equal to 0

$$\hat{m}_{ij} = x_{ij} + R_i \left(\frac{\hat{m}_{ij}}{m_{i+}} \right). \tag{19}$$

Initial estimates of the $\{\hat{m}_{ij}\}$ were considered as

$$m_{ij}^{(0)} = \left(\frac{x_{ij}}{x_{i+}} \right) m_{i+}. \tag{20}$$

where $m_{i+} = x_{i+} + R_i$.

On the first iteration, from (19),

$$\hat{m}_{ij}^{(1)} = x_{ij} + R_i \left(\frac{\hat{m}_{ij}^{(0)}}{m_{i+}} \right). \tag{21}$$

So, on the $(k+1)^{th}$ iteration,

$$\hat{m}_{ij}^{(k+1)} = x_{ij} + R_i \left(\frac{\hat{m}_{ij}^{(k)}}{m_{i+}} \right), \tag{22}$$

when $k \rightarrow \infty$, $|\hat{m}_{ij}^{(k+1)} - \hat{m}_{ij}^{(k)}| \leq \epsilon$.

If an underlying Binomial sampling scheme is assumed, then

$$p_{ij} = \frac{x_{ij}}{x_{i+}}.$$

Therefore, from (20),

$$\hat{p}_{ij}^{(0)} = \hat{m}_{ij}^{(0)} / m_{i+}$$

and

$$\hat{p}^{(0)} = \sum_i \sum_j \hat{m}_{ij}^{(0)} / N.$$

On the $(k+1)^{th}$ iteration,

$$\hat{p}_{ij}^{(k+1)} = \hat{m}_{ij}^{(k+1)} / m_{i+}$$

and

$$\hat{p}^{(k+1)} = \sum_i \sum_j \hat{m}_{ij}^{(k+1)} / N.$$

Table 3: Underlying probabilities for a 2x2 table

Fully Classified Table		Row Supplemental Margin
$(1 - \lambda_{1(1)})\pi_{11}$	$(1 - \lambda_{1(1)})\pi_{12}$	$\lambda_{1(1)}\pi_{1+}$
$(1 - \lambda_{1(2)})\pi_{21}$	$(1 - \lambda_{1(2)})\pi_{22}$	$\lambda_{1(2)}\pi_{2+}$

Formulation of Newton-Raphson

From Le (1992), the iterative solution for a parameter estimation on $(k+1)^{th}$ iteration will be considered as

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \Delta \hat{\theta}, \quad (23)$$

where θ is the parameter and

$$\Delta \hat{\theta} = - \left(\frac{d \ln L}{d \theta} \right) / \left(\frac{d^2 \ln L}{d \theta^2} \right).$$

Differentiating (7) with γ_{ij} and equal to 0, then results in

$$\begin{aligned} m_{ij} = & R_{i+} \left[\frac{m_{ij}}{m_{i+}} - \left(\frac{m_{ij}}{m_{i+}} \right)^2 \right] \\ & + C_{+j} \left[\frac{m_{ij}}{m_{+j}} - \left(\frac{m_{ij}}{m_{+j}} \right)^2 \right] + D \left[\frac{m_{ij}}{m_{++}} - \left(\frac{m_{ij}}{m_{++}} \right)^2 \right] \end{aligned} \quad (24)$$

To avoid the confusion of m_{ij} for (7) and (24), let $m1_{ij}$ and $m2_{ij}$ for (7) and (24), respectively.

For application of the Newton-Raphson method in the two-way incomplete contingency table, consider

$$\hat{m}1_{ij}^{(1)} = x_{ij} + R_{i+} \left(\frac{\hat{m}_{ij}^{(0)}}{m_{i+}^{(0)}} \right) + C_{+j} \left(\frac{\hat{m}_{ij}^{(0)}}{m_{+j}^{(0)}} \right) + D \left(\frac{\hat{m}_{ij}^{(0)}}{m_{++}^{(0)}} \right) \quad (25)$$

$$\begin{aligned} m2_{ij}^{(1)} = & R_{i+} \left[\frac{m1_{ij}^{(0)}}{m1_{i+}^{(0)}} - \left(\frac{m1_{ij}^{(0)}}{m1_{i+}^{(0)}} \right)^2 \right] \\ & + C_{+j} \left[\frac{m1_{ij}^{(0)}}{m1_{+j}^{(0)}} - \left(\frac{m1_{ij}^{(0)}}{m1_{+j}^{(0)}} \right)^2 \right] \\ & + D \left[\frac{m1_{ij}^{(0)}}{m1_{++}^{(0)}} - \left(\frac{m1_{ij}^{(0)}}{m1_{++}^{(0)}} \right)^2 \right] \end{aligned} \quad (26)$$

where $\hat{m}_{ij}^{(0)}$ is the same with (9) and

$$\hat{m}_{ij}^{(1)} = m1_{ij}^{(0)} - \frac{m1_{ij}^{(0)}}{m2_{ij}^{(0)}}. \quad (27)$$

On the $(k+1)^{th}$ iteration,

$$\begin{aligned} \hat{m}1_{ij}^{(k)} = & x_{ij} + R_{i+} \left(\frac{\hat{m}_{ij}^{(k-1)}}{m_{i+}^{(k-1)}} \right) \\ & + C_{+j} \left(\frac{\hat{m}_{ij}^{(k-1)}}{m_{+j}^{(k-1)}} \right) + D \left(\frac{\hat{m}_{ij}^{(k-1)}}{m_{++}^{(k-1)}} \right) \end{aligned} \quad (28)$$

$$\begin{aligned} m2_{ij}^{(k)} = & R_{i+} \left[\frac{m1_{ij}^{(k-1)}}{m1_{i+}^{(k-1)}} - \left(\frac{m1_{ij}^{(k-1)}}{m1_{i+}^{(k-1)}} \right)^2 \right] \\ & + C_{+j} \left[\frac{m1_{ij}^{(k-1)}}{m1_{+j}^{(k-1)}} - \left(\frac{m1_{ij}^{(k-1)}}{m1_{+j}^{(k-1)}} \right)^2 \right] \\ & + D \left[\frac{m1_{ij}^{(k-1)}}{m1_{++}^{(k-1)}} - \left(\frac{m1_{ij}^{(k-1)}}{m1_{++}^{(k-1)}} \right)^2 \right] \end{aligned} \quad (29)$$

$$\hat{m}_{ij}^{(k+1)} = m1_{ij}^{(k)} - \frac{m1_{ij}^{(k)}}{m2_{ij}^{(k)}}. \quad (30)$$

For an accelerated convergence, these equations were employed to obtain the maximum likelihood estimators.

The EM algorithm: Formulation of the EM algorithm for contingency table

The EM approach for incomplete categorical data on the basis of Multinomial, Binomial and Poisson assumptions is now investigated.

Multinomial Distributions

For Multinomial distributions, the complete data log likelihood is

$$\begin{aligned} \log L_c(\pi_i) = & \sum_{i=1}^{n-1} (x_i + z_i) \log \pi_i + (x_n + z_n) \\ & \log(1 - \pi_1 - \pi_2 - \dots - \pi_{n-1}), \end{aligned} \quad (31)$$

where unobservable or missing data are referred to as $z_i = (z_1, z_2, \dots, z_n)^T$ and $z_i = r_i + c_i + d_i$ with r_i being missing column data, c_i missing row data, and d_i both row and column missing data on cell i^{th} . Differentiating (31) with respect to π_i , results in

$$\hat{\pi}_i = \frac{x_i + z_i}{x_n + z_n} \pi_n. \quad (32)$$

Since $\sum_{i=1}^n \pi_i = 1$, therefore from (32),

$$\hat{\pi}_i = \frac{x_i + z_i}{N} \quad (33)$$

where $\sum_{i=1}^n (x_i + z_i) = N$.

The E- and M-values on the first iteration for cell (i, j) were considered as follows.

E-step:

$$m_{ij}^{(1)} = x_{ij} + R_{i+} \left(\frac{\pi_{ij}}{\pi_{i+}} \right) + C_{+j} \left(\frac{\pi_{ij}}{\pi_{+j}} \right) + D\pi_{ij}$$

where $m_{ij}^{(1)}$ is the expected of cell (i, j) on the first iteration.

M-step:

$$\hat{\pi}_{ij}^{(1)} = m_{ij}^{(1)} / N,$$

where $\pi_{ij}^{(1)}$ is the probability for cell (i, j).

On the (k+1)th iteration, the E- and M-steps were defined as follows:

E-step:

$$m_{ij}^{(k+1)} = x_{ij} + R_{i+} \left(\frac{\hat{\pi}_{ij}^{(k)}}{\hat{\pi}_{i+}^{(k)}} \right) + C_{+j} \left(\frac{\hat{\pi}_{ij}^{(k)}}{\hat{\pi}_{+j}^{(k)}} \right) + D\pi_{ij}$$

M-step:

$$\hat{\pi}_{ij}^{(k+1)} = m_{ij}^{(k+1)} / N.$$

The E- and M-steps were alternated and repeated until

$$\begin{aligned} \left| \hat{\pi}_{ij}^{(k+1)} - \hat{\pi}_{ij}^{(k)} \right| &= \left| \frac{m_{ij}^{(k+1)}}{N} - \frac{m_{ij}^{(k)}}{N} \right| \\ &= \left| \frac{m_{ij}^{(k+1)} - m_{ij}^{(k)}}{N} \right| \leq \varepsilon. \end{aligned}$$

Therefore, when $k \rightarrow \infty$, $\lim_{k \rightarrow \infty} \left| \hat{\pi}_{ij}^{(k+1)} - \hat{\pi}_{ij}^{(k)} \right| = 0$, and $\hat{\pi}_{ij}^{(k+1)} = \hat{\pi}_{ij}^{(k)} = \pi^*$.

Binomial distribution

For the binomial distribution, the complete-data log likelihood is

$\log L_c(p_{i1}) = (x_{i1} + z_{i1}) \log p_{i1} + (x_{i2} + z_{i2}) \log(1 - p_{i1})$, for $i = 1, \dots, n$, and z_i is referred to as unobservable or missing data on the i^{th} row where $z_{i1} + z_{i2} = z_i$. Differentiating with respect to p_{i1} results in

$$\hat{p}_{i1} = \frac{x_{i1} + z_{i1}}{x_{i1} + x_{i2} + z_i}. \quad (34)$$

From (34), if all rows are summed, the following is obtained

$$\sum_{i=1}^I \hat{p}_{i1} = \frac{\sum_{i=1}^I r_{i1} + z_{i1}}{\sum_{i=1}^I r_{i1} + r_{i2} + z_i}.$$

Since $\sum_{i=1}^I r_{i1} + r_{i2} + z_i = N$, total sample,

$$\hat{p}_{+j} = \frac{\sum_{i=1}^n x_{ij}^{(1)}}{N}.$$

The E- and M-values on the first iteration for cell (i, j) were considered as follows:

E-step:

$$m_{ij}^{(1)} = x_{ij} + R_{i+} p_{ij}$$

where $m_{ij}^{(1)}$ is the expected value of cell (i, j) on the first iteration and $p_{ij} = x_{ij}/x_{i+}$.

M-step:

$$\hat{p}_{ij}^{(1)} = m_{ij}^{(1)} / (m_{i1}^{(1)} + m_{i2}^{(1)}),$$

and

$$\hat{p}_{+j}^{(1)} = \frac{\sum_{i=1}^n m_{ij}^{(1)}}{N}$$

On the (k+1)th iteration, the E- and M-steps were defined as follows:

E-step:

$$m_{ij}^{(k+1)} = x_{ij} + R_{i+} p_{ij}^{(k)}$$

M-step:

$$\hat{p}_{ij}^{(k+1)} = m_{ij}^{(k+1)} / (m_{i1}^{(k+1)} + m_{i2}^{(k+1)}),$$

and

$$\hat{p}_{+j}^{(k+1)} = \frac{\sum_{i=1}^n m_{ij}^{(k+1)}}{N}$$

The E- and M-steps were alternated and repeated until

$$\left| \hat{p}_{ij}^{(k+1)} - \hat{p}_{ij}^{(k)} \right| = \left| \frac{m_{ij}^{(k+1)}}{m_{i1}^{(k+1)} + m_{i2}^{(k+1)}} - \frac{m_{ij}^{(k)}}{m_{i1}^{(k)} + m_{i2}^{(k)}} \right| \leq \varepsilon,$$

and

$$\left| \hat{p}_{+j}^{(k+1)} - \hat{p}_{+j}^{(k)} \right| = \left| \frac{\sum_{i=1}^n m_{ij}^{(k+1)}}{N} - \frac{\sum_{i=1}^n m_{ij}^{(k)}}{N} \right| = \left| \frac{\sum_{i=1}^n m_{ij}^{(k+1)} - \sum_{i=1}^n m_{ij}^{(k)}}{N} \right| \leq \varepsilon.$$

Therefore, when $k \rightarrow \infty$, $\lim_{k \rightarrow \infty} \left| \hat{p}_{ij}^{(k+1)} - \hat{p}_{ij}^{(k)} \right| = 0$

and $\lim_{k \rightarrow \infty} \left| \hat{p}_{+j}^{(k+1)} - \hat{p}_{+j}^{(k)} \right| = 0$.

Poisson distribution

For the Poisson distribution, the complete-data log likelihood is

Log $L_c(y; \theta_i) =$

$$\sum_{i=1}^n [(x_i + z_i) \log(\theta_i) - \theta_i - \log(x_i + z_i)!] \tag{35}$$

where $z_1 + z_2 + \dots + z_n$ is referred to as unobservable or missing data. By differentiating (35) with respect to θ_i ,

$$\hat{\theta}_i = x_i + z_i. \tag{36}$$

Referring to Figure 1, the E- and M-values on the first iteration for the cell (i, j) was considered as:

E-step:

$$z_{ij}^{(1)} = R_{ij}^{(1)} + C_{ij}^{(1)} + D_{ij}^{(1)},$$

where

$$R_{ij}^{(1)} = R_{i+} \left(\frac{x_{ij}}{x_{i+}} \right), C_{ij}^{(1)} = C_{+j} \left(\frac{x_{ij}}{x_{+j}} \right), \text{ and}$$

$$D_{ij}^{(1)} = D \left(\frac{x_{ij}}{x} \right).$$

M-step:

$$\hat{\theta}_{ij}^{(1)} = x_{ij} + z_{ij}^{(1)}.$$

On the (k+1)th iteration, the E- and M-steps were defined as follows:

E-step:

$$z_{ij}^{(k+1)} = R_{ij}^{(k+1)} + C_{ij}^{(k+1)} + D_{ij}^{(k+1)},$$

where

$$R_{ij}^{(k+1)} = R_{i+} \left(\frac{\theta_{ij}^k}{\theta_{i+}^k} \right), C_{ij}^{(k+1)} = C_{+j} \left(\frac{\theta_{ij}^k}{\theta_{+j}^k} \right), \text{ and}$$

$$D_{ij}^{(k+1)} = D \left(\frac{\theta_{ij}^k}{N} \right),$$

and N is total sample.

M-step:

$$\hat{\theta}_{ij}^{(1)} = x_{ij} + z_{ij}^{(1)}.$$

The E- and M-steps were alternated and repeated until

$$\left| \hat{\theta}_{ij}^{(k+1)} - \hat{\theta}_{ij}^{(k)} \right| = \left| (x_{ij} + z_{ij}^{(k+1)}) - (x_{ij} + z_{ij}^{(k)}) \right| = \left| z_{ij}^{(k+1)} - z_{ij}^{(k)} \right| \leq \varepsilon$$

Therefore, when $k \rightarrow \infty$

$$\lim_{k \rightarrow \infty} \left| \hat{\theta}_{ij}^{(k+1)} - \hat{\theta}_{ij}^{(k)} \right| = 0,$$

and it may be said that $\hat{\theta}_{ij}^{(k+1)} = \hat{\theta}_{ij}^{(k)} = \theta^*$.

Results

The results of MLE, adopting Newton-Raphson in MLE and the M-step of the EM algorithm for the Poisson distribution are presented in Tables 4, 5 and 6 respectively.

Table 4 MLE for Poisson distribution

Iteration	Cells			
	(1,1)	(1,2)	(2,1)	(2,2)
1	17.99	10.13	19.23	11.66
2	17.72	10.05	19.35	11.89
3	17.67	10.03	19.35	11.96
4	17.66	10.02	19.34	11.98
5	17.67	10.02	19.33	11.99
6	17.67	10.01	19.33	11.99
7	17.67	10.01	19.33	11.99

Table 5: Adopting Newton-Raphson in MLE for Poisson distribution

Iteration	Cells			
	(1,1)	(1,2)	(2,1)	(2,2)
1	17.92	9.7	19.62	11.76
2	17.68	9.38	19.87	12.06
3	17.66	9.34	19.78	12.22
4	17.65	9.34	19.77	12.24
5	17.65	9.34	19.77	12.24

Table 6: M-step for Poisson distribution

Iteration	Cells			
	(1,1)	(1,2)	(2,1)	(2,2)
1	17.99	10.13	19.22	11.66
2	17.72	10.04	19.35	11.9
3	17.67	10.02	19.35	11.96
4	17.67	10.02	19.34	11.98
5	17.67	10.01	19.33	11.98
6	17.67	10.01	19.33	11.99
7	17.67	10.01	19.33	11.99

The results of MLE, adopting Newton-Raphson in MLE and the M-step of the EM algorithm for the Multinomial distribution are presented in Tables 7, 8 and 9 respectively. The results of MLE and the M-step for the Binomial distribution are presented in Tables 10, 11 and 12 respectively.

Based upon results, both the MLE and

the EM algorithms converge on the 7th iteration (see Tables 4 and 6), and both methods give the same results. However, by adopting the Newton-Raphson in the MLE, the results on the 5th iteration were obtained (see Table 5). Although it seems that the EM algorithm was converging the same as the MLE, the EM algorithm involves two calculation steps on each iteration.

In other words, the EM takes longer to compute the results compared with the MLE. After adopting the Newton-Raphson in the MLE, it was able to give faster convergence without as much deviance in the results as the EM algorithm. Tables 7 and 8 were obtained by considering the last iteration of Tables 4 and 5 respectively. The results were also the same for the Multinomial distribution for the MLE and the EM algorithm. By comparing the results of Table 8 with the last iteration of Table 9, it is observed that the results are not much different. However Table 11 was obtained by considering the last iteration of Table 10. Results shown in Tables 11 and 12 were the same as those obtained for the Binomial distribution.

Testing independence

For two-way contingency tables, the null hypothesis of statistical independence is $H_0 : \pi_{ij} = \pi_{i+} \pi_{+j}$ for all i and j . The likelihood-ratio statistic, G^2 is asymptotically equivalent to χ^2 when $n \rightarrow \infty$ with d.f. = $(r - 1)(c - 1)$ where r is the number of rows and c is the number of columns in the contingency table.

According to Schafer (1997), $G^2 = 2[\ell(\hat{\pi}|Y_{obs}) - \ell(\tilde{\pi}|Y_{obs})]$, where $\ell(\hat{\pi}|Y_{obs})$ is the unrestricted ML estimate ($\hat{\pi}$) and $\ell(\tilde{\pi}|Y_{obs})$ is the restricted ML estimate ($\tilde{\pi}$). Thus,

$\ell(\pi|Y_{obs})$ is considered as:

$$\ell(\pi|Y_{obs}) = \ell_A(\pi|Y_{obs}) + \ell_B(\pi|Y_{obs}) + \ell_C(\pi|Y_{obs}) + \ell_D(\pi|Y_{obs}).$$

For the Multinomial and Poisson distributions with the MLE and EM algorithm,

$$G^2 = 2[\ell(\hat{\pi}|Y_{obs}) - \ell(\tilde{\pi}|Y_{obs})].$$

For the Binomial distribution, $H_0: p = p_{i1}$, therefore

$$G^2 = 2[\ell(\hat{p}|Y_{obs}) - \ell(\tilde{p}|Y_{obs})],$$

where $\ell(\hat{p}|Y_{obs})$ is the unrestricted ML estimate of \hat{p} and $\ell(\tilde{p}|Y_{obs})$ is the restricted ML estimate of \tilde{p} . For both the MLE and the EM algorithms $\ell(p|Y_{obs})$ is considered as:

$$\ell(p|Y_{obs}) = \ell_A(p|Y_{obs}) + \ell_B(p|Y_{obs}).$$

Therefore, adopting the Newton-Raphson in the MLE and EM algorithms for Multinomial and Poisson distributions, $G^2=0.02$. However, for the Binomial distribution, $G^2=0.01$. From these results, it may be concluded that treatment type is independent of the results of treatment for the Multinomial and Poisson distributions, and the number of seizure pain which is less than five is the same for treatment 0 and 1 for the Binomial distribution.

Table 7: MLE for the Multinomial distribution

Cells			
(1,1)	(1,2)	(2,1)	(2,2)
0.2992	0.1697	0.3276	0.2032

Table 8: Adopting Newton-Raphson in MLE for the Multinomial distribution

Cells			
(1,1)	(1,2)	(2,1)	(2,2)
0.2992	0.1583	0.3351	0.2075

Table 9: M-step for Multinomial distribution

Iteration	Cells			
	(1,1)	(1,2)	(2,1)	(2,2)
1	0.3049	0.1716	0.3259	0.1976
2	0.3003	0.1702	0.3278	0.2017
3	0.2995	0.1698	0.3278	0.2027
4	0.2993	0.1698	0.3278	0.2031
5	0.2992	0.1697	0.3276	0.2031
6	0.2992	0.1697	0.3276	0.2032
7	0.2992	0.1697	0.3276	0.2032

Table 10: MLE for the Poisson distribution

Iteration	Cells			
	(1,1)	(1,2)	(2,1)	(2,2)
1	14.95	8.05	16.42	9.58
2	14.95	8.05	16.42	9.58

Table 11: MLE for the Binomial distribution

Cells					
(1,1)	(2,1)	(+, 1)	(1, 2)	(2,2)	(+, 2)
0.65	0.6315	0.6402	0.35	0.3685	0.3598

Table 12: M-step for the Binomial distribution

Iteration	Cells					
	(1,1)	(2,1)	(+, 1)	(1, 2)	(2,2)	(+, 2)
1	0.65	0.6315	0.6402	0.35	0.3685	0.3598
2	0.65	0.6315	0.6402	0.35	0.3685	0.3598

Conclusion

The EM algorithm is more complicated than the MLE, because the EM algorithm involves the E-(expectation) and M-(maximization) steps. This makes the calculations more complicated and also increases the amount of time required to calculate results as compared with the MLE, which is more straightforward for estimating cell probabilities in cases of incomplete categorical data. For example when consider a contingency table with a Poisson sampling scheme, for MLE, the expected value is obtained as in (11) by considering the previous iteration of the expected value. However, for the EM algorithm, before calculating the expected value in the M-step, the E-step - which involves the estimation of initial cell probability first – must first be considered. For the Binomial sampling scheme, the convergence for estimation of p_{i1} and p can be obtained when first considering Poisson sampling employing the MLE procedure. Again, if the EM algorithm is considered, the E-step is required first in order to obtain an initial estimate for p_{ij} . Similar explanations may be given for Multinomial sampling cases where, if MLE is considered, the Poisson sampling must be addressed before using the last iteration to

obtain $\hat{\pi}_{ij}$. The EM algorithm, however, requires step by step convergence starting from the initial value for $\hat{\pi}_{ij}$ before convergence is achieved.

The MLE can better perform by adopting the Newton-Raphson method, because this method helps to accelerate the convergence. When the MLE is adopted with that of Newton-Raphson, as a convergence method, it is clear that the MLE and the EM algorithm are two different kinds of algorithms. The MLE algorithm provides a direct way to maximize the final expected value, while the EM algorithm involves expectation before the maximization; however, the EM algorithm demonstrates the distribution of missing values at each step until convergence on the basis of the marginal probabilities.

The MLE is much simpler than the EM algorithm when one is interested simply in final results. If interest lies in understanding the distribution of missing values in more detail, the EM algorithm is the better choice.

References

- Baker, S. G. (1994). Composite linear models for incomplete multinomial data. *Statist. Med.*, 13, 609-622.
- Baker, S. G., & Laird, N. M. (1988). Regression analysis for categorical variables with outcome subject to nonignorable nonresponse. *Journal of the American Statistical Association*, 83, 62-69.
- Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press: Cambridge, MA.
- Chen, T., & Fienberg, S. E. (1974). Two-dimensional contingency tables with both completely and partially cross-classified data. *Biometrics*, 30, 629-642.
- Chen, T., & Fienberg, S. E. (1976). The analysis of contingency tables with incompletely classified data. *Biometrics*, 32, 133-144.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. B*, 39, 1-38.
- Diggle, P. J., Liang K. Y., & Zeger S. L. (1994) *Analysis of Longitudinal Data*. Oxford University Press: Oxford, England.
- Fay, R. E. (1986). Causal models for patterns of nonresponse. *Journal of the American Statistical Association*, 81, 354-365.
- Fienberg, S. E. (1970). Quasi-independence and maximum likelihood estimation in incomplete contingency tables. *Journal of the American Statistical Association*, 65, 1610-1616.
- Fienberg, S. E. (1980). *The analysis of cross-classified categorical data*. MIT Press: Cambridge, MA.
- Fuchs, C. (1982). Maximum likelihood estimation and model selection in contingency tables with missing data. *Journal of the American Statistical Association*, 77, 270-278.
- Galecki, A. T., Have, T. R. T. & Molenberghs, G. (2001). A simple and fast alternative to the EM algorithm for incomplete categorical data and latent class models. *Comp. Stat. & Data Analysis*, 35, 265-281.
- Lauritzen, S. L. (1995). The EM algorithm for graphical association models with missing data. *Comp. Stat. & Data Analysis*, 19, 191-201.
- Lauritzen, S. L., & Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society B*, 50, 157-224.
- Le, C. T. (1992). *Fundamentals of Biostatistical Inference*. Marcel Dekker, Inc.: New York, NY.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. John Wiley: New York, NY.
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*, 2nd Ed. John Wiley: New Jersey.
- Lyles, R. H., & Allen, A. S. (2003). Missing data in the 2x2 table: patterns and likelihood-based analysis for cross-sectional studies with supplemental sampling. *Stats. Med.*, 22, 517-534.
- McLachlan, G. J., & Krishnan, T. (1997). *The EM algorithm and extensions*. John Wiley: New York, NY.
- Molenberghs, G., & Goetchebeur, E. (1997). Simple fitting algorithm for incomplete categorical data. *Journal of the Royal Statistical Society B*, 59, 401-414.
- Nordheim, E. V. (1984). Inference from nonrandomly missing categorical data: an example from a genetic study on Turner's syndrome. *Journal of the American Statistical Association*, 79, 772-780.
- Phillips, M. J. (1993). Contingency tables with missing data, *The Statistician*, 42, 9-18.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Chapman and Hall/CRC: New York, NY.
- Stasny, E. A. (1985). Modeling non-ignorable non-response in panel data. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 349-354.
- Tang, M. L., Ng, K. W., Tian, G. L. & Tan, M. (2007). On improved EM algorithm and confidence interval construction for incomplete rxc tables. *Computational Statistics and Data Analysis*, 51: 2919-2933.