

## MLE for Incomplete Contingency Tables with Lagrangian Multiplier

Shin-Soo Kang<sup>1)</sup>

### Abstract

Maximum likelihood estimate(MLE) is obtained from the partial log-likelihood function for the cell probabilities of two way incomplete contingency tables proposed by Chen and Fienberg(1974). The partial log-likelihood function is modified by adding lagrangian multiplier that constraints can be incorporated with. Variances of MLE estimators of population proportions are derived from the matrix of second derivatives of the loglikelihood with respect to cell probabilities.

Simulation results, when data are missing at random, reveal that Complete-case(CC) analysis produces biased estimates of joint probabilities under MAR and less efficient than either MLE or MI. MLE and MI provides consistent results under either the MAR situation. MLE provides more efficient estimates of population proportions than either multiple imputation(MI) based on data augmentation or complete case analysis. The standard errors of MLE from the proposed method using lagrangian multiplier are valid and have less variation than the standard errors from MI and CC.

**Keywords** : Complete case analysis, Multiple imputation, Partial log-likelihood

### 1. Introduction

In the analysis of contingency tables, it may happen that some observations are not fully observed. This issue relating incomplete contingency tables has been studied for a long time. One simple approach, known as complete-case(CC) analysis, discards the missing data by restricting analysis to only fully classified counts in an incomplete contingency table.

---

1) Professor, Department of Management and Information, Kwandong University, Kangnung, 210-701, Korea  
E-mail: sskang@kd.ac.kr

Chen and Fienberg (1974) used an iterative procedure for computing maximum likelihood estimates and developed Pearson and likelihood ratio tests of independence for two-way tables for which either the row classification or the column classification could be missing for some cases. As in Chen and Fienberg (1974), Hocking and Oxspring (1974) consider three independent multinomial distributions corresponding to the set of fully cross-classified counts and the two sets of partially classified counts, where either the row classification or the column classification is missing.

An alternative approach involves constructing a complete table, in which all cases are completed classified, by imputing information for the missing row or column classification. Multiple imputation, proposed by Rubin (1978), provides a way to take advantage of commonly used tests of independence for completely classified tables.

West et al.(2002) proposed a method using an EM algorithm and a Bayesian prior to analyze longitudinal studies with repeated measures. They generate all possible sets of missing values, form a set of possible complete data sets, and then weight each data set according to clearly defined assumptions. They apply an appropriate statistical test procedure to each data set, combining the results to give an overall indication of significance.

Instead of imputing information for missing classifications, maximum likelihood estimates of population proportions can be obtained from the observed information, including both the completely and partially classified cases. Little(1982) used a simple EM algorithm to estimate cell probabilities.

Maximum likelihood estimates of population proportions is obtained from the partial log-likelihood function for the cell probabilities of two way incomplete contingency tables proposed by Chen and Fienberg(1974). This MLE method does work under MAR. Constraints such as sum of cell probabilities is equal to one can be incorporated with lagrangian multiplier proposed by Aitchison and Silvey(1958). Variances of MLE estimators of population proportions are derived from the matrix of second derivatives of the loglikelihood with respect to cell probabilities.

The performances of MLE, MI and the CC approach are examined through the Monte Carlo studies in section 4. The bias and efficiency of estimates of population proportions are considered.

## 2. Notation and Likelihood Function

Consider an  $I \times J$  contingency table where the row factor  $X_1$  has  $I$  categories and the column factor  $X_2$  has  $J$  categories. Assume simple random

sampling with replacement. In a complete table, where the row and column categories are observed for every case in the sample, the counts have a multinomial distribution with sample size  $N$  and probability vector  $\Theta$  where  $\Theta = (\Theta_{11}, \Theta_{12}, \dots, \Theta_{1j}, \Theta_{21}, \dots, \dots, \Theta_{IJ})$ . Let  $\Theta_{ij}$  be an element of  $\Theta$  denote the population proportion for the cell  $(i, j)$ .

When information on either the row or column classification is missing, we can construct a table of counts for the completely classified cases where  $x_{ij}$  denotes the number of cases observed in the  $(i, j)$  cell. We can also construct one-way tables of counts for partially classified cases. Let  $x_{im}$  denote the number of cases in the  $i^{\text{th}}$  row category,  $i = 1, 2, \dots, I$ , where the column category is unknown, and let  $x_{mj}$  denote the number of cases in the  $j^{\text{th}}$  column category,  $j = 1, 2, \dots, J$ , where the row category is unknown. Then,  $x_{im}$  and  $x_{mj}$  are marginally observed counts on a single variable. Let  $x_{mm}$  denote the number of cases where both the row and column categories are missing. The total sample size is

$$\begin{aligned} N &= \sum_{ij} x_{ij} + \sum_i x_{im} + \sum_j x_{mj} + x_{mm} \\ &= x_{cc} + x_{+m} + x_{m+} + x_{mm}. \end{aligned}$$

The partial log-likelihood function for the cell probabilities  $\Theta$  proposed by Chen and Fienberg(1974) is the following;

$$l(\Theta) = \sum_i \sum_j x_{ij} \log \Theta_{ij} + \sum_j x_{mj} \log \Theta_{+j} + \sum_i x_{im} \log \Theta_{i+} \quad (1)$$

The log-likelihood function in (1) does not include  $x_{mm}$ . Discarding the  $x_{mm}$  cases for which both variables are missing does not affect any results in this paper except  $N$  changed to  $n = N - x_{mm}$  because those cases do not contain any information about the joint distribution of  $X_1$  and  $X_2$ .

### 3. MLE of Cell Probabilities and Its Variance

The EM algorithm can be used to get a MLE. Little(1982) illustrates EM procedure with partially classified data on two variables. The EM algorithm is quite simple. The estimates at  $t^{\text{th}}$  iteration is

$$\Theta_{ij}^{(t)} = 1/n \left( x_{ij} + x_{im} \times \frac{\Theta_{ij}^{(t-1)}}{\Theta_{i+}^{(t-1)}} + x_{mj} \times \frac{\Theta_{ij}^{(t-1)}}{\Theta_{+j}^{(t-1)}} \right), \text{ where } \Theta_{ij}^{(0)} = \frac{x_{ij}}{x_{cc}}. \text{ The estimates at}$$

the first step of EM procedure. We can get a MLE of  $\theta$  after convergence.

For the example by Little(1982), the 30 partially classified units with  $X_1=1$  have  $X_2=1$  with probability  $100/(100+50)$  and  $X_2=2$  with probability  $50/(100+50)$ . Thus in effect,  $(30)(100)/150 = 20$  are allocated to  $X_2=1$  and  $(30)(50)/150 = 10$  are allocated to  $X_2=2$  at the first step. In the next step new cell probabilities are computed from the completed data and the procedure iterates to convergence.

Since we have a constraint such as  $\sum_{ij} \theta_{ij} = 1$ , the likelihood function (1) can be expressed with lagrangian multiplier as

$$l(\theta^*) = \sum_r \sum_j x_{ij} \log \theta_{ij} + \sum_r x_{mj} \log \theta_{+j} + \sum_r x_{im} \log \theta_{i+} + \gamma (1 - \sum_{ij} \theta_{ij}),$$

where  $\theta^* = (\theta', \gamma)'$ . The first derivative of  $l(\theta^*)$  with respect to  $\theta_{ij}$  and  $\gamma$  is

$$\begin{aligned} \frac{\partial l(\theta^*)}{\partial \theta_{ij}} &= \frac{x_{ij}}{\theta_{ij}} + \frac{x_{mj}}{\theta_{+j}} + \frac{x_{im}}{\theta_{i+}} - \gamma \\ \frac{\partial l(\theta^*)}{\partial \gamma} &= 1 - \sum_{ij} \theta_{ij} \end{aligned}$$

The second partial derivatives are

$$\begin{aligned} \frac{\partial^2 l(\theta^*)}{\partial \theta_{ij}^2} &= \frac{-x_{ij}}{\theta_{ij}^2} - \frac{x_{mj}}{\theta_{+j}^2} - \frac{x_{im}}{\theta_{i+}^2}, \\ \frac{\partial^2 l(\theta^*)}{\partial \theta_{ij} \partial \theta_{is}} &= -\frac{x_{im}}{\theta_{i+}^2}, \quad \frac{\partial^2 l(\theta^*)}{\partial \theta_{ij} \partial \theta_{kj}} = -\frac{x_{mj}}{\theta_{+j}^2} \\ \frac{\partial^2 l(\theta^*)}{\partial \theta_{ij} \partial \theta_{ks}} &= 0, \quad \frac{\partial^2 l(\theta^*)}{\partial \theta_{ij} \partial \gamma} = -1, \quad \frac{\partial^2 l(\theta^*)}{\partial \gamma^2} = 0 \end{aligned}$$

Let  $l(\theta^*)$  be the matrix of second derivatives of the loglikelihood with respect to  $\theta^*$ . An estimate of the covariance matrix of  $\widehat{\theta}$  which is a MLE of  $\theta^*$  is  $-\Gamma^{-1}(\widehat{\theta}^*)$ . Let  $\widehat{\Sigma}_M$  be a matrix omitting the last row and column of  $-\Gamma^{-1}(\widehat{\theta}^*)$ .  $\widehat{\Sigma}_M$  gives us an estimate of the covariance matrix of  $\widehat{\theta}_M$ , where  $\widehat{\theta}_M$  is a MLE of  $\theta$  computing through EM procedure. For  $2 \times 2$  table,  $\theta^* = (\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}, \gamma)'$  and  $l(\theta^*)$  is

$$I(\Theta^*) = - \begin{pmatrix} d_{11} & \frac{X_{1m}}{\Theta_{2+}^2} & \frac{X_{m1}}{\Theta_{2+}^2} & 0 & 1 \\ \frac{X_{1m}}{\Theta_{2+}^2} & d_{12} & 0 & \frac{X_{m2}}{\Theta_{2+}^2} & 1 \\ \frac{X_{m1}}{\Theta_{2+}^2} & 0 & d_{21} & \frac{X_{2m}}{\Theta_{2+}^2} & 1 \\ 0 & \frac{X_{m2}}{\Theta_{2+}^2} & \frac{X_{2m}}{\Theta_{2+}^2} & d_{22} & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix},$$

where  $d_{ij} = \frac{X_{ij}}{\Theta_{ij}^2} + \frac{X_{mj}}{\Theta_{+j}^2} + \frac{X_{im}}{\Theta_{i+}^2}$ .

We can get a MLE from other algorithms like Newton-Raphson using the matrix of second derivatives of the loglikelihood with respect to  $\Theta^*$  instead of EM algorithm, but both methods provide similar results of MLE.

#### 4. Simulation Results

The performances of MLE, MI and the CC approach are examined through the Monte Carlo studies. All of the  $2 \times 2$  incomplete contingency tables for this study were generated with equal probability margins and sample size  $N=500$  such that  $(\Theta_{11}, \Theta_{12}, \Theta_{21}, \Theta_{22}) = (0.2, 0.3, 0.3, 0.2)$ . There are two cases with different missing at random mechanism (MAR) and 1000 tables were generated for each case. The missing mechanism for each case is

$$\begin{aligned} \Pr(X_1 \text{ is missing} | X_2 = 1) &= m_1 \\ \Pr(X_1 \text{ is missing} | X_2 = 0) &= m_2 \\ \Pr(X_2 \text{ is missing}) &= m_3 \end{aligned} \quad (2)$$

For the first case  $m_1=0.1, m_2=0.3, m_3=0.2$  and for the second case  $m_1=0.2, m_2=0.4, m_3=0.3$  in (2), respectively. The percentages of cases with missing information on at least one variable are expected to be 36% and 51% for case 1 and 2 respectively. We will compare three methods to check point estimations of  $\Theta_{11}$  and  $\Theta_{1+}\Theta_{+1}-\Theta_{11}$  and standard errors of  $\widehat{\Theta}_{11}$ . The algorithms for MI were programmed through S-PLUS 6.1(2001) functions for missing values.

Table 1 shows means and standard deviations of 1000 values for the estimates of  $\Theta_{11}$  from the generated 1000 tables in this section. The true value of  $\Theta_{11}$  is

0.2. MLE and MI methods provide essentially unbiased estimates for the cell probabilities but CC does not. The standard deviations of the estimates differ across methods. Complete-case analysis provides the estimate of  $\theta_{11}$  with the largest variance. For all methods, variation increases as the proportion of missing values increases. MLE tends to provide smaller standard deviations of cell proportion than MI.

<Table 1> Point estimation of  $\theta_{11}$

Case	MLE		MI		CC	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	0.1984	0.02063	0.1988	0.02118	0.2242	0.02377
2	0.1978	0.02193	0.1977	0.02281	0.2281	0.02697

Table 2 shows means and standard deviations of 1000 simulated values for the estimates of  $\theta_{1+}\theta_{+1}-\theta_{11}$ , a measure of association between the two variables, when the true value of  $\theta_{1+}\theta_{+1}-\theta_{11}$  is 0.05. The averages of the estimates are similar for all methods. The complete-case and MLE have similar standard deviations and they exhibit smaller standard deviations than MI.

<Table 2> Point estimation of  $\theta_{1+}\theta_{+1}-\theta_{11}$

Case	MLE		MI		CC	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	0.0495	0.01376	0.0493	0.01422	0.0488	0.01358
2	0.0498	0.01608	0.0496	0.01691	0.0490	0.01585

Table 3 shows means and standard deviations of 1000 standard errors of  $\hat{\theta}_{11}$ . The averages of the standard errors are close to standard deviations of  $\theta_{11}$  in Table 1 for all methods. It says that all three methods provides valid standard errors. The standard errors of MLE obtained from the matrix of second derivatives of the loglikelihood have less variation than the standard errors from MI and CC. MI standard errors have the largest variation among three methods.

<Table 3> Standard errors of  $\hat{\theta}_{11}$ 

Case	MLE		MI		CC	
	Mean	S.D	Mean	S.D.	Mean	S.D.
1	0.0202	0.00080	0.0206	0.00188	0.0233	0.00099
2	0.0222	0.00094	0.0227	0.00322	0.0268	0.00128

## 5. Conclusion

It has been an issue to estimate the variance of MLE for the cell probabilities of an incomplete contingency table because it is very complicated to get the second derivatives of the likelihood. The likelihood with lagrangian multiplier can solve this problem. The standard errors of MLE obtained from the matrix of second derivatives of the loglikelihood are valid and have less variation than the standard errors from MI and CC.

Complete-case(CC) analysis produces biased estimates of joint probabilities under MAR and less efficient than either MLE or MI. MLE and MI provides consistent results under either the MAR situation.

When data are missing at random, simulation results reveal that MLE provides more efficient estimates of population proportions than either multiple imputation(MI) based on data augmentation or complete case analysis, but neither MLE nor MI provides an improvement over complete-case(CC) analysis with respect to  $\theta_{1+} \theta_{+1} - \theta_{11}$ , which is a measure of association between the two variables.

## References

1. Aitchison, J., and Silvey, S. D. (1958). Maximum-Likelihood Estimation of Parameters Subject to Restraints, *The annals of Mathematical Statistics*, 29, 813-828.
2. Chen, T. T., and Fienberg, S. E. (1974). Two-dimensional contingency tables with both completely and partially cross-classified data, *Biometrics*, 32, 133-144.
3. Hocking, R. R., and Oxspring, H. H. (1974). The Analysis of Partially Categorized contingency Data, *Biometrics*, 30, 469-483.
4. Little, R. J. A., and Rubin, D. B. (2002). *Statistical analysis with missing data*, J. Wiley and Sons, New York.
5. Rubin, D. B. (1978). Multiple Imputation in Sample Surveys - A Phenomenological Bayesian Approach to Nonresponse, *Proceedings of the Survey Research Methods Section, American Statistical Association*, 1978, 20-34.

6. S-Plus 6.1 Manual: Analyzing Data with Missing Values in S-Plus (2001). Insightful Corporation. Seattle, Washington.
7. West, C. P., and Dawson, J. D. (2002). Complete imputation of missing repeated categorical data: one-sample applications, *Statistics in Medicine*, 21, 203-217.

[ received date : Jun. 2006, accepted date : Jul. 2006 ]