Journal of the Korean Data & Information Science Society
2015, **26**(4), 971–984

http://dx.doi.org/10.7465/jkdi.2015.26.4.971
한국데이터정보과학회지

Large tests of independence in incomplete two-way contingency tables using fractional imputation

Shin-Soo Kang[1] · Michael D. Larsen[2]

[1]Department of MIS, Catholic Kwandong University
[2]Department of Statistics, George Washington University

Received 10 January 2015, revised 21 April 2015, accepted 19 May 2015

Abstract

Imputation procedures fill-in missing values, thereby enabling complete data analyses. Fully efficient fractional imputation (FEFI) and multiple imputation (MI) create multiple versions of the missing observations, thereby reflecting uncertainty about their true values. Methods have been described for hypothesis testing with multiple imputation. Fractional imputation assigns weights to the observed data to compensate for missing values. The focus of this article is the development of tests of independence using FEFI for partially classified two-way contingency tables. Wald and deviance tests of independence under FEFI are proposed. Simulations are used to compare type I error rates and Power. The partially observed marginal information is useful for estimating the joint distribution of cell probabilities, but it is not useful for testing association. FEFI compares favorably to other methods in simulations.

Keywords: Delta method, hot deck, maximum likelihood estimation, missing data, multiple imputation, wald statistic.

1. Introduction

Consider a two-dimensional contingency table defined by row factor X_1 having I categories and column factor X_2 having J categories. Assume data are gathered using simple random sampling with replacement. In a complete table, the counts have a multinomial distribution with sample size N and probability vector θ. Let n_{ij} denote the count in cell (i,j) with θ_{ij}, an element of θ, the corresponding population proportion.

When information on either the row or column classification is missing, one can construct a table of counts for the completely classified cases. Let x_{ij} denotes the counts observed in cell (i,j). One can also construct one-way tables for partially classified cases. Let x_{im} denote the number of cases with unknown column category in the i^{th} row category, $i = 1,2,\ldots,I$ and x_{mj} denote the number of cases with unknown row category in the j^{th} column category, $j = 1,2,\ldots,J$. There could be a number of cases (x_{mm}) with row and column categories both missing. The total sample size is $N = \sum_{ij} x_{ij} + \sum_i x_{im} + \sum_j x_{mj} + x_{mm}$ $= x_{cc} + x_{\bullet m} + x_{m\bullet} + x_{mm}$.

[1] Professor, Department of MIS, Catholic Kwandong University, Kangneung 210-701, Korea.
[2] Corresponding author: Department of Statistics, George Washington University, U.S.A.
E-mail: mlarsen@bsc.gwu.edu

1.1. Literature review and existing methods

Complete case (CC) analysis discards the observations missing any information and restricts analysis to only fully classified counts. Such an analysis usually assumes that the x_{cc} observations are randomly chosen from a multinomial distribution with parameters θ. The maximum likelihood estimates of the proprtions then are $\hat{\theta}_{ij} = x_{ij}/x_{cc}$. Loss of sample size can be substantial. If the data are missing because of the category they would have been in, then tests of independence can be biased (Little and Rubin, 2002; Magder, 2003).

Chen and Fienberg (1974) used an iterative procedure for computing maximum likelihood estimates of population proportions from the observed information, including both the completely and partially classified cases. They developed Pearson and likelihood ratio tests of independence for two-way tables for which either the row or column classification could be missing for some cases. Kang (2006) proposed the likelihood function with lagrangian mutiplier to get maximum likelihood estimates of population proportions and its variance for incomplete contingency tables.

Little (1982; examples 3.4 and 3.5) used the EM algorithm (Dempster *et al.*, 1977) to estimate cell probabilities in incomplete contingency tables via maximum likelihood estimation (MLE). Little (1982; example 4.3) also used the EM algorithm for MLE when categorical variables are missing not at random (NMAR), in the sense of Rubin (1976). Fuchs (1982) fit log linear models to incomplete tables and demonstrated procedures for testing model fit. Lipsitz *et al.* (1998) show how to use generalized linear model software for MLE of cell probabilities using the connection between the multinomial and Poisson likelihoods. Friendly (2000) described a reformulation of the problem for the general case of missing data in a multiway table and implemented this algorithm as a SAS macro program.

Variance estimation after MLE can be based on the inverse of the observed Fisher information or via a simulation technique, such as the Jackknife or Bootstrap (e.g., Little and Rubin, 2002; chapter 5). If the EM algorithm is used to produce the estimates, then the SEM algorithm (Meng and Rubin, 1991) can be used for variance estimation. For some contingency table models, the ECM algorithm may be used for estimation (Meng and Rubin, 1993) and variance calculation (Van Dyk *et al.,* 1995). A score test for independence in two-way incomplete contingency tables was proposed by Lipsitz and Fitzmaurice (1996).

Simple imputation methods fill-in the missing data and pretend that the values are real. It is well-known that these procedures lead to overstatement of precision and possible bias (e.g., Little and Rubin, 2002; chapter 3). Imputing the most common category or the most common category conditional on some observed values would probably distort relationships with other variables. It generally is based on a very strong assumption, such as all who do not report outcome data in smoking cession programs are smoking. Instead, one might randomly impute missing values for a unit according to the distribution of the similar respondents. The random draws could be taken from the empirical distribution by randomly sampling donors (called a hot-deck procedure) or from the conditional multinomial distribution estimated based on the complete cases. Despite the stochastic nature of the imputation, pretending that imputed values are real could lead to overstatement of statistical significance.

Multiple imputation, proposed by Rubin (1978) and described in Rubin (1987, 1996), replaces each missing value with a set of draws from a posterior predictive distribution in order to capture the variability due to the missing data. Schafer (1997; chapter 7) provides details of algorithms for categorical data with missing values. For each completed data set, composed of the observed data plus one imputation for each missing value, one computes parameter estimates and standard errors. These are combined to produce a parameter and total variability estimate. The total variablity reflects the between and within imputation variability. It is possible to obtain a single test statistic by combining statistics from multiply imputed data sets. Li *et al.* (1991) proposed a Wald test statistic. Meng and Rubin (1992) proposed likelihood ratio tests with F reference distributions.

Fractional imputation (FI) uses multiple responding units as donors for a missing unit. Each donor is asigned a weight, like a sampling weight. FI was discussed originally by Kalton and Kish (1981) and later by Fay (1996). Kim and Fuller (2004) describe fully efficient fractional imputation (FEFI), which uses every responding unit within a designated imputation group as a donor for a unit with missing information. Each imputed value is assigned a weight. Fractional imputation enables consistent variance estimates assuming the data are missing at random (MAR) conditional on the observed information and the missing data are ignorable (Rubin, 1976). Fay (1996) and Kim and Fuller (2004) describe the algorithms for estimation of a mean. Kang, Koehler, and Larsen (2011) present procedures for FEFI for incomplete two-way contingency tables.

The literature on Bayesian methods for categorical data is not discussed here. Bayesian methods for categorical data often are applied to situations with repeated observations over time (West and Dawson, 2002), variable selection (Green and Park, 2003), graphical models (Geng *et al.*, 2000; Geng *et al.*, 2003), and hierarchical models (Wakefield, 2004). These areas as well as work on multivariate categorical variables with arbitrary patterns of missing data will be considered in future work on FI and FEFI.

1.2. Outline and plan

Section 2 describes FEFI for partially classified two-way tables and reviews other methods. Four numerical examples that were included in Kang *et al.* (2011) are summarized. Section 3 reviews methods of Kang *et al.* (2011) for obtaining covariance matrices for FEFI estimates of population proportions and variance estimation for other methods. Section 4 proposes Wald and deviance tests of independence using the results of Kang *et al.* (2011) when data are missing completely at random (MCAR). Reference F distributions with adjusted denominator degrees of freedom are suggested. Type I error levels and Power of the FEFI and other methods are studied through simulation in Section 5. Section 6 is a summary and discussion.

2. Point estimation

2.1. Fully efficient fractional imputation under MCAR

As described in Kang *et al.* (2011), the FEFI imputation group for a unit with observed value of the $X_1 = x_1$ but missing X_2 is the set of complete cases with the same value of $X_1 = x_1$. Imputation fractions for the J possible values of X_2 are obtained from the conditional frequencies of X_2 in the cross-classified table of complete cases given the observed value of X_1. Similarly, the imputation group for a unit with a missing value of X_1, but an observed value $X_2 = x_2$ is the set of complete cases with the value of $X_2 = x_2$ with imputation fractions determined by the I conditional frequencies of X_1 in the complete cases. For an observation with both variables missing, the imputation group is the set of all remaining cases. In this case, imputation fractions correspond to the joint frequencies of X_1 and X_2 incorporating all partial information.

The counts in the completed table obtained by FEFI under simple random sampling are given by the following formula:

$$\begin{aligned} \hat{n}^*_{ij} &= x_{ij} + x_{ij}\left(\frac{x_{im}}{x_{i\bullet}} + \frac{x_{mj}}{x_{\bullet j}}\right) + \frac{x_{ij}x_{mm}}{N - x_{mm}}\left(1 + \frac{x_{im}}{x_{i\bullet}} + \frac{x_{mj}}{x_{\bullet j}}\right) \\ &= x_{ij}\left(1 + \frac{x_{im}}{x_{i\bullet}} + \frac{x_{mj}}{x_{\bullet j}}\right)\left(1 + \frac{x_{mm}}{N - x_{mm}}\right), \end{aligned} \tag{2.1}$$

where x_{ij} is the observed count for fully observed cases prior to imputation, $x_{i\bullet} = \sum_{j=1}^{J} x_{ij}$ and $x_{\bullet j} = \sum_{i=1}^{I} x_{ij}$. The total sample size is $N = \sum_{ij} \hat{n}^*_{ij}$.

Discarding the x_{mm} cases that clearly provide no information about cell or marginal proportions, (2.1) can be simplified as

$$\hat{n}_{ij} = x_{ij}\left(1 + \frac{x_{im}}{x_{i\bullet}} + \frac{x_{mj}}{x_{\bullet j}}\right), \tag{2.2}$$

with N changed to $n = N - x_{mm}$. Doing so does not affect the FEFI estimates of the population proportions:

$$\begin{aligned} \hat{\theta}_{FEFI} &= \frac{1}{n}(\hat{n}_{11}, \hat{n}_{12}, \cdots, \hat{n}_{1J}, \hat{n}_{21}, \hat{n}_{22}, \cdots, \hat{n}_{2J}, \cdots, \hat{n}_{IJ})' \\ &= \frac{1}{N}(\hat{n}^*_{11}, \hat{n}^*_{12}, \cdots, \hat{n}^*_{1J}, \hat{n}^*_{21}, \hat{n}^*_{22}, \cdots, \hat{n}^*_{2J}, \cdots, \hat{n}^*_{IJ})'. \end{aligned} \tag{2.3}$$

2.2. Other methods

Complete case analysis can use maximum likelihood estimation based on only the observations with both X_1 and X_2 observed. The EM algorithm can be used to produce maximum likelihood estimates of parameters under a multinomial model with completely and partially classified units. Single stochastic imputation randomly picks a donor with complete information for each partially observed case. Unlike FEFI, this method uses only the complete cases for imputation.

For multiple imputation (MI) assuming a multinomial model for the data, one can choose a Jeffreys noninformative prior distribution (Box and Tiao, 1992) on the parameters. The prior distribution on θ is taken to be proportional to $(\prod_{ij} \theta_{ij})^{-1/2}$. If all counts had been fully observed, the posterior distribution of θ would have been a Dirichlet distribution with parameter values $x_{ij} + 1/2$. Data augmentation (e.g., Schafer, 1997; chapter 7) is used to generate imputations for each missing value. Let $\hat{\theta}_d$, $d = 1, \ldots, D$ be the standard complete-data estimates for the vector of cell probabilities from the D imputed data sets. The MI point estimates of the cell probabilities are given by

$$\bar{\theta}_D = \frac{1}{D}\sum_{d=1}^{D} \hat{\theta}_d. \tag{2.4}$$

2.3. Numerical examples

The imputation procedures are illustrated with four numerical examples that were described in Kang *et al.* (2011). Table 2.1 shows data from a simple random sample of size $N = 88$ where X_1 and X_2 assume values of 0 or 1 and the symbol ? is used to indicate a missing value. MCAR is assumed. Observed counts for the 9 possible response patterns in line 3 with the resulting FEFI counts in line 4.

Table 2.1 Response patterns for a 2×2 incomplete contingency table with FEFI computations.

X_1	1	1	0	0	?	?	1	0	?
X_2	1	0	1	0	1	0	?	?	?
Original counts	5	10	15	20	8	9	6	7	8
FEFI counts	9.9	18.7	26.5	33					

Magder (2003) reports the outomce (success or failure) of two medical procedures for treatment of nonarteritic anterior ischemic optic neuropathy. The two methods are optic nerve decompression surgery (ONDS) and careful follow-up (Careful). Table 2.2 presents observed counts and the resulting FEFI table.

Table 2.2 Observed data from Magder (2003) and FEFI table.

X_1, ONDS vs. Careful	1	1	0	0	1	0
X_2, Success vs. Failure	1	0	1	0	?	?
Counts	37	70	47	67	20	17
FEFI table	43.9	83.1	54	77		

Chen and Fienberg (1974) present data on 456 premature infants cross-classified by a "health" index and a measurement of serum bilirium level. Table 2.3 presents observed counts and the resulting FEFI table for this example.

Table 2.3 Response patterns for a 2×2 incomplete contingency table with FEFI computations for data from Chen and Fienberg (1974). For serum reading, 1 indicates a level of 0-1.0 and 0 indicates a level of greater than 1.0. For the "health" index, 1 indicates a score 0-6 and 0 a score 7-10.

Serum	1	1	0	0	?	?	1	0
Health	1	0	1	0	1	0	?	?
Original counts	35	75	57	112	117	36	11	13
FEFI counts	83	96.9	133.9	142.2				

Rubin *et al.* (1995) report the results of the Slovenian Public Opinion Survey (Table 2.4). Survey respondents were asked about planning at attend a upcoming plebiscite and their support for independence. Clustering of survey units is ignored in this presentation. Future work will address the use of the cluster structure, additional variables, and hypotheses about the missingness mechansims, such as those of Molenberghs, Kenward, and Goetghebeur (2001).

Table 2.4 Response patterns for a 2×2 incomplete contingency table with FEFI imputations, 1=yes and 0=no.

Attend	1	1	0	0	?	?	1	0	?
Independence	1	0	1	0	1	0	?	?	?
Original counts	1439	78	16	16	144	54	159	32	136
FEFI counts	1853.8	140.2	36.0	44.1					

Point estimates for the four examples and a Monte Carlo comparison of point estimators was presented in Kang *et al.* (2011).

3. Variances of point estimators

Variance estimation was discussed in Kang *et al.* (2011) and is reviewed briefly here. Let $C_0 = (x_{11}, \ldots, x_{1J}, x_{21}, \ldots, x_{2J}, \ldots, x_{IJ}, x_{m1}, \ldots, x_{mJ}, x_{1m}, \ldots x_{Im})'$, and $q = I \times J + I + J$. Conditional on the value of x_{mm}, C_0 has a multinomial distribution with sample size $n = N - x_{mm}$ and probabilities

$$\pi = (\pi_{11}, \cdots, \pi_{1J}, \pi_{21}, \cdots, \pi_{2J}, \cdots \pi_{IJ}, \pi_{m1}, \cdots, \pi_{mJ}, \pi_{1m}, \cdots \pi_{Im})'. \tag{3.1}$$

The variance-covariance matrix of C_0 is

$$Var(C_0) = n\,(\Delta_\pi - \pi\pi')\,, \tag{3.2}$$

where Δ_π is a diagonal matrix with the elements of π on the main diagonal.

Each element of $\hat{\theta}_{FEFI}$ is a function of the elements of C_0. Using the delta method, the variance of $\hat{\theta}_{FEFI}$ is

$$Var(\hat{\theta}_{FEFI}) = \frac{1}{n} D\,(\Delta_\pi - \pi\pi')\,D' \equiv \Sigma_F, \tag{3.3}$$

where $p = I \times J$, and

$$D_{p\times q} = \begin{pmatrix} \frac{\partial \hat{n}_{11}}{\partial x_{11}} & \cdots & \frac{\partial \hat{n}_{11}}{\partial x_{IJ}} & \frac{\partial \hat{n}_{11}}{\partial x_{m1}} & \cdots & \frac{\partial \hat{n}_{11}}{\partial x_{mJ}} & \frac{\partial \hat{n}_{11}}{\partial x_{1m}} & \cdots & \frac{\partial \hat{n}_{11}}{\partial x_{Im}} \\ \frac{\partial \hat{n}_{12}}{\partial x_{11}} & \cdots & \frac{\partial \hat{n}_{12}}{\partial x_{IJ}} & \frac{\partial \hat{n}_{12}}{\partial x_{m1}} & \cdots & \frac{\partial \hat{n}_{12}}{\partial x_{mJ}} & \frac{\partial \hat{n}_{12}}{\partial x_{1m}} & \cdots & \frac{\partial \hat{n}_{12}}{\partial x_{Im}} \\ \vdots & \vdots & & & & & & & \vdots \\ \frac{\partial \hat{n}_{IJ}}{\partial x_{11}} & \cdots & \frac{\partial \hat{n}_{IJ}}{\partial x_{IJ}} & \frac{\partial \hat{n}_{IJ}}{\partial x_{m1}} & \cdots & \frac{\partial \hat{n}_{IJ}}{\partial x_{mJ}} & \frac{\partial \hat{n}_{IJ}}{\partial x_{1m}} & \cdots & \frac{\partial \hat{n}_{IJ}}{\partial x_{Im}} \end{pmatrix},$$

with

$$\frac{\partial \hat{n}_{ij}}{\partial x_{cd}} = \begin{cases} 1 + \left(\frac{x_{im}}{x_{i\bullet}} + \frac{x_{mj}}{x_{\bullet j}}\right) - x_{ij}\left(\frac{x_{im}}{x_{i\bullet}^2} + \frac{x_{mj}}{x_{\bullet j}^2}\right), & c = i \text{ and } d = j \\ -\left(\frac{x_{ij}x_{im}}{x_{i\bullet}^2}\right) & , \; c = i,\; d \neq j,\; \text{and } d \neq m \\ -\left(\frac{x_{ij}x_{mj}}{x_{\bullet j}^2}\right) & , \; c \neq i,\; d = j,\; \text{and } c \neq m \\ \frac{x_{ij}}{x_{\bullet j}} & , \; c = m \text{ and } d = j \\ \frac{x_{ij}}{x_{i\bullet}} & , \; c = i \text{ and } d = m \\ 0 & , \; \text{otherwise} \end{cases}$$

By the CLT, the FEFI estimators have an approximate multivariate normal distribution with mean θ and variance Σ_F. The variance estimator substitutes $\hat{\pi} = C_0/n$ into formula 3.3.

Variance estimation for the other estimators was discussed in Kang *et al.* (2011). Kang *et al.* (2011) examined standard errors in the four examples and reported a Monte Carlo simulation of standard error estimation and confidence interval coverage.

4. Tests of independence

The null hypothesis of statistical independence is

$$H_0 : \theta_{ij} = \theta_{i\bullet}\theta_{\bullet j}, \quad \text{for all } i \text{ and } j. \tag{4.1}$$

Defining

$$g_{ab}(\theta) \equiv \left(\sum_{j=1}^{J} \theta_{aj}\right)\left(\sum_{i=1}^{I} \theta_{ib}\right) - \theta_{ab}$$

the null hypothesis of statistical independence can be expressed as

$$H_0 : g_{ab}(\theta) = 0 \text{ for all } a = 1, 2, \ldots, I \text{ and } b = 1, 2, \ldots, J. \tag{4.2}$$

Let $g(\theta) = (g_{11}(\theta), \cdots, g_{1J}(\theta), g_{21}(\theta), \cdots, g_{2J}(\theta), \cdots, g_{IJ}(\theta))'$, then (4.2) is equivalent to $H_0 : g(\theta) = \mathbf{0}$.

4.1. Tests in a complete contingency table

For a complete two-dimensional contingency table with sample size N, the estimator of θ is $\hat{\theta}$, where $\hat{\theta_{ij}} = x_{ij}/N$. By the Central Limit Theorem, it has an approximate multivariate normal distribution with variance $\Sigma = V_\theta/N = \Sigma(\theta)$, where $V_\theta = (\Delta_\theta - \theta\theta')$ and Δ_θ is a diagonal matrix with the elements of θ on the main diagonal. Under H_0, $g(\hat{\theta})$ has an approximate $p = IJ$ dimensional normal distribution with mean $\mathbf{0}$ and variance $G\Sigma G'$, where $G_{p\times p} = G(\theta)$ is the matrix of first partial derivatives of $g(\theta)$ with respect to θ_{ij}. The elements of $G_{p\times p}$ are

$$\frac{\partial g_{ab}(\theta)}{\partial \theta_{ij}} = \begin{cases} \sum_{i=1}^{I} \theta_{ib} + \sum_{j=1}^{J} \theta_{aj} - 1, & \text{for } a = i, \text{ and } b = j \\ \sum_{i=1}^{I} \theta_{ib} & , \text{ for } a = i, \text{ and } b \neq j \\ \sum_{j=1}^{J} \theta_{aj} & , \text{ for } a \neq i, \text{ and } b = j \\ 0 & , \text{ for } a \neq i, \text{ and } b \neq j \end{cases}.$$

A Wald statistic for testing H_0 is

$$\hat{Q} = g(\hat{\theta})'\hat{T}^{-}g(\hat{\theta}), \tag{4.3}$$

where $\hat{T} = (\hat{G}\hat{\Sigma}\hat{G}')$ is obtained by substituting $\hat{\theta}$ for θ (Shao, 1999; pages 386-387) and the superscript $-$ denotes generalized inverse (Pringle and Rayner, 1971). For a complete table, $\hat{Q}$ has a distribution converging to a central chi-squared with $df = k = (I-1)(J-1)$ when H_0 is true. A more accurate F approximation for the small sample distribution of $\hat{Q}$ is

$$\hat{Q}\frac{N-k}{k(N-1)} \approx F(k, N-k). \tag{4.4}$$

For complete tables, the likelihood ratio test or deviance statistic, $2\sum_{i,j} \hat{n}_{ij} \log(\hat{n}_{ij}/\hat{m}_{ij})$, where $\hat{m}_{ij} = \hat{n}_{i\bullet}\hat{n}_{\bullet j}/n$ is obtained from the complete table, is asymptotically distributed as $\chi^2(k)$ under the null hypothesis, and is asymptotically equivalent to the Wald statistic. The Pearson χ^2 statistic, $\sum_{i,j}(\hat{n}_{ij} - \hat{m}_{ij})^2/\hat{m}_{ij}$, is asymptotically equivalent to the deviance test statistic as well.

4.2. Tests of independence using FEFI estimates in an incomplete contingency table

Under H_0, $g(\hat{\theta}_{FEFI})$ has an approximate p dimensional normal distribution with variance $G\Sigma_F G'$ and expectation zero, where Σ_F is defined by (3.3). A Wald statistic H_0 is

$$\hat{Q}_F = g(\hat{\theta}_{FEFI})'\hat{T}_F^- g(\hat{\theta}_{FEFI}), \tag{4.5}$$

where $\hat{T}_F = (\hat{G}_F\hat{\Sigma}_F\hat{G}_F')$ and $\hat{G}_F$ is obtained by substituting $\hat{\theta}_{FEFI}$ for θ in $G(\theta)$, $\hat{\Sigma}_F$ is obtained by substituting $x_{i\bullet}x_{\bullet j}/x_{cc}$ for x_{ij} in V_θ, and π is evaluated at the sample proportion, C_0/n, in (3.3).

The sample size, N, in (4.4) should be adjusted for the levels of missingness of information on row and column categories in an incomplete contingency table. In a complete table, $\text{Var}(\hat{\theta}) = V_\theta/N$ and $\text{Var}\Big(g(\hat{\theta})\Big) = GV_\theta G'/N$ by the delta method. The adjusted sample size n^* can be chosen as

$$n^* = \text{tr}(GV_\theta G'\hat{T}_F^-)/k = \text{tr}(n\hat{T}_{CF}\hat{T}_F^-)/k, \tag{4.6}$$

where $\hat{T}_{CF}$ is an estimate of $\text{Var}\Big(g(\hat{\theta})\Big)$ based on the FEFI completed table considered as the fully observed table without missing values with sample size $n = N - x_{mm}$ and $\hat{T}_{CF}$ is obtained under the assumption that H_0 is true. Then we can apply the test to an $I \times J$ incomplete two-way contingency table by rejecting the null hypothesis of independence when

$$\hat{Q}_F \frac{n^* - k}{k(n^* - 1)} > F(k, n^* - k, \alpha), \tag{4.7}$$

where α is the Type I error level of the test.

An approximate Wald statistic that is a multiple of the familiar Pearson statistic can be derived by replacing $\hat{T}_F^-$ with $r\hat{T}_{CF}^-$ in (4.5), where $r \equiv \text{tr}(\hat{T}_{CF}\hat{T}_F^-)/k$ is a scalar summary of the difference between the covariance matrices for a table with no missing information and an FEFI completed table. This yields

$$\begin{aligned} Q_F^* &= r \times g(\hat{\theta}_{FEFI})'\hat{T}_{CF}^- g(\hat{\theta}_{FEFI}) \\ &= r \sum_{all\ cells} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}, \end{aligned} \tag{4.8}$$

which is a multiple of usual Pearson χ^2 statistic calculated from the FEFI completed table. We can use the F reference distribution in (4.7) substituting Q_F^* for $\hat{Q}_F$.

For incomplete tables, the Pearson χ^2 statistic in (4.8) is asymptotically equivalent to the deviance test statistic, $2\sum_{i,j}\hat{n}_{ij}\log(\hat{n}_{ij}/\hat{m}_{ij})$, where $\hat{m}_{ij} = \hat{n}_{i\bullet}\hat{n}_{\bullet j}/n$ is obtained from the FEFI completed table. So an approximate deviance test statistic computed from the FEFI completed table is

$$D_F \equiv 2r\frac{n^* - k}{k(n^* - 1)}\sum_{i,j}\hat{n}_{ij}\log(\hat{n}_{ij}/\hat{m}_{ij}), \tag{4.9}$$

for which the null hypothesis is rejected if $D_F > F(k, n^* - k, \alpha)$.

4.3. Tests based on other methods of handling missing data

For complete case (CC) analysis and single stochastic imputation (SSI), tests of independence are implemented as they would presumably be implemented in practice. That is, methods for complete tables are applied to the CC table or the SSI table as if they were the fully observed and true tables. The CC method under MCAR should have the desired type I error levels, but less power than other methods. SSI should produce inflated type I errors, due to the underestimation of variance.

Tests for independence when maximum likelihood estimation (MLE) is used to produce parameter estimates naturally are based on likelihood (deviance) or Pearson tests. As demonstrated by Chen and Fienberg (1974), these goodness-of-fit tests each consist of three parts:

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(x_{ij}-\hat{a}_{ij})^2}{\hat{a}_{ij}} + \sum_{i=1}^{I}\frac{(x_{im}-\hat{r}_i)^2}{\hat{r}_i} + \sum_{j=1}^{J}\frac{(x_{mj}-\hat{c}_j)^2}{\hat{c}_j} \tag{4.10}$$

and

$$G^2 = 2\left[\sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij}\log\frac{x_{ij}}{\hat{a}_{ij}} + \sum_{i=1}^{I} x_{im}\log\frac{x_{im}}{\hat{r}_i} + \sum_{j=1}^{J} x_{mj}\log\frac{x_{mj}}{\hat{c}_j}\right], \tag{4.11}$$

where

$$\begin{aligned} \hat{a}_{ij} &= x_{cc}\left(\frac{x_{i\bullet}+x_{im}}{x_{cc}+x_{+m}}\right)\left(\frac{x_{\bullet j}+x_{mj}}{x_{cc}+x_{m+}}\right), \\ \hat{r}_i &= x_{+m}\left(\frac{x_{i\bullet}+x_{im}}{x_{cc}+x_{+m}}\right), \\ \hat{c}_j &= x_{m+}\left(\frac{x_{\bullet j}+x_{mj}}{x_{cc}+x_{m+}}\right). \end{aligned}$$

The test statistics in (4.10) and (4.11) use $IJ-1$ instead of $(I-1)(J-1)$ degrees of freedom, because the margins also are included in the test statistic. Procedures based on maximum likelihood should have their desired type I error levels and be more powerful than CC. As seen in an example, the MLE procedures are also sensitive to the MCAR assumption; these MLE tests can lead to rejecting independence when independence might be true and sampling is MAR. Extensions of the Chen and Fienberg (1974) work for Wald tests and for applications with some observations totally unclassified will be the focus of future work.

Multiple imputation (MI) produces overall estimates of parameters and corresponding sampling variances. Li *et al.* (1991) introduced a modified Wald test procedure. Since it is easy in these examples to create many more imputations than the dimension of the parameters, it is possible to conduct a Wald test based on the MI estimates of parameters and the MI estimate of the variance-covariace matrix for the estimates; see, for example, Meng and Rubin (1992). Meng and Rubin (1992) introduced a likelihood ratio based procedure and applied it to a three-way partially classified table of counts. Li *et al.* (1991) introduced a procedure for combining results from either Wald test statistics or LRT statistics. Since the Pearson test statistic is asymptotically equivalent to these two test statistics for complete two-way tables, the same procedure should be applicable for combining results from Pearson tests. In applications in this paper, Li *et al.* (1991)'s procedure, presented in their Section

2, is applied to the Pearson chi-squared test of independence. In all cases, the number of imputations (m) is taken to be 5. Schafer (1997) also describes these procedures, except the application to the Pearson test statistic, in his Section 4.3.3.

The first procedure requires the overall estimate of the parameter, the average of the within-imputation variance estimates, and the between imputation variance estimate. The third procedure requires only the test statistics from appling the test procedure separately to the multiply imputed tables. As such, it is expected that the third procedure will be less powerful than the others (Meng and Rubin, 1992). The second procedure requires a level of information between the other two. We do not know of an application in the literature of the third procedure to Pearson tests of independence in two-way tables. Note that the dimension of θ in a two-by-two table is one under the independence assumption, so that a small number of imputation sets should be sufficient to implement the above procedures.

Table 4.1 presents results of applying tests of independence to the four examples. Single stochastic imputation (SSI) produces larger test statistics and smaller p-values than the complete case (CC) method. This reflects the understatement of uncertainty using SSI. Maximum likelihod (MLE), multiple imputation (MI), and fully efficient fractional imputation (FEFI) produce larger statistics than CC, but similar p-values. These three methods utilize partially classified information, whereas CC does not. Example 4 produces some seemingly erratic results, because expected sample sizes in some cells in example 4 are much smaller than others. Example 3 using MLE produces a result at odds with the other methods. Chen and Fienberg (1974)'s test statistic captures the fact that, apparently, the data are not missing completely at random, but rather missing at random depending on the observed marginal classification. Data in the power simulation will be MCAR, so this situation will not arise. The existence of such an example, however, has important implications for inference in general, and will be examined carefully in future work.

Table 4.1 Test statistics (*Stat*) and p-values (*pval*) for testing independence in four examples using five methods: complete cases (CC), maximum likelihood estimation (MLE), single stochastic imputation (SSI), multiple imptuation (MI), and fully efficient fractional imputation (FEFI).

		Likelihood/ G^2 deviance test		Pearson X^2 statistic		Wald test statistic	
		Stat	pval	Stat	pval	Stat	pval
Example 1	CC	.40	.53	.40	.53	.40	.53
Artificial	MLE	1.83	.61	1.82	.61	–	–
	SSI	.77	.38	.76	.38	.76	.38
	MI	.46	.57	.71	.44	.47	.56
	FEFI	.43	.52	.42	.52	.44	.51
Example 2	CC	1.04	.31	1.04	.31	1.04	.31
Magder	MLE	1.44	.49	1.44	.49	–	–
(2003)	SSI	1.21	.27	1.21	.27	1.21	.27
	MI	1.05	.32	1.05	.32	1.05	.32
	FEFI	1.04	.31	1.04	.31	1.04	.31
Example 3	CC	.11	.74	.11	.74	.11	.74
Chen and	MLE	78.18	.00	75.30	.00	–	–
Fienberg	SSI	.24	.62	.24	.62	.24	.62
(1974)	MI	.17	.75	.47	.52	.17	.75
	FEFI	.13	.72	.13	.72	.13	.72
Example 4	CC	49.71	.00	110.63	.00	110.63	.00
Rubin *et al.*	MLE	186.55	.00	255.83	.00	–	–
(1995)	SSI	116.26	.00	216.63	.00	216.63	.00
	MI	71.84	.00	74.89	.00	56.53	.00
	FEFI	59.85	.00	110.28	.00	20.49	.00

5. Monte carlo results to compare five methods for 2 × 2 Tables

The performance of tests of independence described in the previous section for the five imputation methods are examined through Monte Carlo simulations. Type I error levels are estimated from 10,000 tables simulated under the independence assumption. Power levels are examined by simulating 10,000 tables under an alternative to independence.

5.1. Type I error levels

Two-by-two tables were generated to estimate Type I error levels under an independence model with some observations randomly seleceted to be missing. Ten thousand tables were generated under each of four scenarios. In all cases X_1 and X_2 were independently generated as Bernoulli(0.5) and Bernoulli(0.7) random variables, respectively, so that $\theta = (0.35, 0.15, 0.35, 0.15)$, $\theta_{1\bullet}\theta_{\bullet 1} - \theta_1 1 = 0$. The four scenarios are defined by two factors. The first factor is sample size with levels, $n = 200$ or $n = 400$. The second factor is the missingness mechanism. For $k = 1, \ldots, n$, let M_{vk} have a value of 1 if X_{vk} is missing and 0 otherwise, where $v = 1, 2$. Either M_{vk}, $v = 1, 2$ are both Bernoulli(0.3) or M_{1k} is Bernoulli(0.3) and M_{2k} is Bernoulli(0.6) independently for all $k = 1, \ldots, n$ and $v = 1, 2$ and independent of X_{vk}.

Table 5.1 shows the numbers of tables for which the independence null hypothesis was falsely rejected out of 10000 simulated tables for three nominal Type I error levels. All methods, expect single stochastic imputation (SSI), have reasonable Type I error levels. SSI rejects the true null hypotheses much too often. Type I error levels can be inflated for the MI method in cases with a large proportion of missing values.

Table 5.1 Type I errors out of 10,000 data sets for four scenarios using five methods at three significance levels: complete cases (CC), maximum likelihood estimation (MLE), single stochastic imputation (SSI), multiple imptuation (MI), and fully efficient fractional imputation (FEFI). Number of expected rejections is 100 at 1%, 500 at 5%, and 1000 at 10% significance.

		G^2			X^2			Wald		
P(miss)		1%	5%	10%	1%	5%	10%	1%	5%	10%
n	Method	100	500	1000	100	500	1000	100	500	1000
(.3, .3)	CC	93	515	1034	90	495	1026	73	495	999
$n = 200$	MLE	93	493	1013	81	457	968	–	–	–
	SSI	633	1731	2548	623	1719	2537	599	1693	2516
	MI	107	514	1019	138	590	1082	186	695	1222
	FEFI	87	503	1027	80	484	1015	103	531	1052
(.3, .3)	CC	96	529	1021	93	526	1015	86	510	995
$n = 400$	MLE	100	494	1016	94	486	1003	–	–	–
	SSI	712	1647	2451	707	1639	2451	696	1628	2436
	MI	121	532	1006	139	559	1043	193	687	1175
	FEFI	92	523	1012	88	518	1005	99	533	1027
(.3, .6)	CC	121	552	1071	106	513	1041	78	456	985
$n = 200$	MLE	134	535	1040	104	496	982	–	–	–
	SSI	1193	2312	3138	1175	2304	3125	1140	2276	3104
	MI	119	555	1052	224	731	1312	339	933	1457
	FEFI	111	524	1042	98	492	1015	142	590	1096
(.3, .6)	CC	109	502	1029	101	494	1018	94	469	991
$n = 400$	MLE	115	526	1002	102	507	978	–	–	–
	SSI	1175	2332	3134	1167	2326	3131	1156	2315	3114
	MI	133	536	1033	165	626	1247	296	882	1419
	FEFI	103	492	1010	96	478	996	124	523	1036

5.2. Power Study

Tables were generated from multinomial distributions with proportions given in equation (5.1). The expected proportions under independence in each cell are 0.25.

$$(\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}) = (0.2, 0.3, 0.3, 0.2) \tag{5.1}$$

Four combinations of missing value probabilities and sample sizes were applied to the simulated tables. The missing indicator variables and the multinomial variables were generated independently in order to satisfy the assumption of MCAR. Combinations one, two, and three had sample sizes of 200, whereas combination four had sample size 400. The probability of being missing in the four samples were 0.1, 0.2, 0.3, and 0.7, respectively.

The numbers in Table 5.2 indicate the number of tables out of 10000 for which the independence null hypothesis was rejected under the given α levels among 10000 tables in each type.

Table 5.2 shows FEFI and CC have more power than other methods. FEFI does not show much improvement on the power levels of the tests of independence over complete-case analysis. Although FEFI is often more conservative than MI in most cases with respect to Type I error levels, the test using the FEFI completed table exhibited more power than MI in all cases. Chen and Fienberg tests are more conservative than FEFI and MI and have less power. As the proportion of missing cases increases from 19% to 91%, the power decreases as expected for all types. SSI has much larger power than the other methods, but it had unreasonably high Type I error levels as well. The increase in power for FEFI using the Wald test could be largely attributed to the increase in the Type I error level for this procedure.

The last two parts in (4.10), for partially classified counts, do not provide much information for the association between the two categorical variables. Those two parts, for partially classified counts, are close to central Chi-square distributions even though the null hypothesis of independence is not true. The test statistic in (4.10) uses $I * J - 1$ instead of $(I-1)(J-1)$ degrees of freedom without providing much of an increase in the value of the test statistic. For these reasons, Chen and Fienberg (1974) tests have less power than other methods.

Table 5.2 Power out of 10,000 data sets for four scenarios using five methods at three levels of significance: complete cases (CC), maximum likelihood estimation (MLE), single stochastic imputation (SSI), multiple imptuation (MI), and fully efficient fractional imputation (FEFI). At 50%, 70% and 90% power there should be 5000, 7000, and 9000 rejections, respectively.

P(miss)		G^2			X^2			Wald		
n	Method	1%	5%	10%	1%	5%	10%	1%	5%	10%
P1: 0.1	CC	5039	7281	8191	4994	7258	8185	4870	7211	8167
$n = 200$	MLE	3342	5639	6836	3241	5578	6800	–	–	–
	SSI	5947	7793	8468	5929	7793	8467	5874	7745	8462
	MI	4562	6961	7968	4646	6985	7981	4796	7071	8037
	FEFI	4909	7229	8169	4873	7207	8165	5093	7294	8194
P2: 0.2	CC	3788	6218	7349	3745	6198	7343	3611	6122	7312
$n = 200$	MLE	2411	4563	5765	2329	4505	5736	–	–	–
	SSI	5709	7503	8207	5687	7502	8198	5634	7464	8184
	MI	2977	5509	6788	3208	5646	6850	3694	5987	7074
	FEFI	3662	6141	7308	3626	6123	7299	3853	6246	7350
P3: 0.3	CC	2887	5204	6438	2831	5187	6425	2650	5099	6382
$n = 200$	MLE	1620	3713	4996	1547	3643	4925	–	–	–
	SSI	5816	7422	8098	5802	7418	8097	5744	7384	8089
	MI	1953	4246	5601	2305	4481	5772	2977	5035	6178
	FEFI	2695	5108	6387	2657	5089	6379	2957	5239	6447
P4: 0.7	CC	833	2242	3308	770	2176	3256	585	1955	3067
$n = 400$	MLE	484	1465	2411	394	1355	2291	–	–	–
	SSI	7349	8152	8523	7348	8149	8520	7330	8138	8516
	MI	352	1436	2385	808	1996	3127	1723	2846	3683
	FEFI	689	2026	3137	615	1950	3097	940	2296	3329

6. Summary and discussion

This paper continues the development of fully efficient fractional imputation (FEFI) for two-way contingency tables with partially classified observed. In Kang *et al.* (2011), variance estimates for FEFI using the delta method were derived and implemented. Simulated coverage levels for methods compared are acceptable for all methods except for SSI, which produces the smallest average standard error. FEFI did not require one to be able to program the EM algorithm or Bayesian data augmentation for computations. It also produces a single estimate, as opposed to MI and SSI, which produce different estimates for different random number seeds. In simulations reported here, there is no evidence that imputation is advantageous for hypothesis testing. Complete case analysis appears to be just as good when the data are missing completely at random, although estimation of parameters θ_{ij} is improved by FEFI.

Since the FEFI estimate is more efficient and stable than MI, tests for independence using FEFI have better power and more appropriate type I error levels. The Chen and Fienberg test yielded the most conservative type I error levels and the smallest power among the three methods because it includes quadratic forms focused on marginal distributions that contain little information about association.

Future work will consider FEFI for categorical data that are missing at random (MAR) and not missing at random (NMAR). It will also consider FEFI for partially classified multivariate categorical data with arbitrary patterns of missing data. In higher dimensions, partially classified observations could be valuable to both estimation and testing and imputation methods could potentially show improvement over complete case analysis.

References

Box, G. E. P. and Tiao, G. C. (1992). *Bayesian inference in statistical analysis (Wiley Classics Library Edition)*, John Wiley & Sons, New York.

Chen, T. 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 B*, **39**, 1-38.

Fay, R. E. (1996). Alternative paradigms for the analysis of imputed survey data *Journal of the American Statistical Association*, **91**, 490-498.

Friendly, M. (2000). Note on "Obtaining the maximum likelihood estimates in incomplete $R \times C$ contingency tables using a Poisson generalized model". *Journal of Computational and Graphical Statistics*, **9**, 158-166.

Fuchs, C. (1982). Maximum likelihood estimation and model selection in contingency tables with missing data. *Journal of the American Statistical Association*, **77**, 270-278.

Geng, Z., He, Y. B., Wang, X. L. and Zhao, Q. (2003). Bayesian method for learning graphical models with incompletely categorical data. *Computational Statistics* & *Data Analysis*, **44**, 175-192.

Geng, Z., Wan, K. and Tao, F. (2000). Mixed graphical models with missing data and the partial imputation EM algorithm. *Scandinavian Journal of Statistics*, **27**, 433-444.

Green, P. E. and Park, T. (2003). A Bayesian hierarchical model for categorical data with nonignorable nonresponse. *Biometrics*, **59**, 886-896.

Kalton, G. and Kish, L. (1981). Two efficient random imputation procedures. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 146-151.

Kang, S. S. (2006). MLE for incompletecontingency tables with lagrangian multiplier. *Journal of the Korean Data & information Science Society*, **17**, 919-925.

Kang, S. S., Koehler, K. J. and Larsen, M. D. (2011). Fractional imputation for incomplete two-way contingency tables. *Metrika*, **75**, 581-599.

Kim, J. K. and Fuller, W. (2004). Fractional hot deck imputation. *Biometrika*, **91**, 559-578.

Li, K. H., Meng, X. L., Raghunathan, T. E. and Rubin, D. (1991). Significance levels from repeated p-values with multiply-imputed data. *Statistica Sinica*, **1**, 65-92.

Li, K. H., Raghunathan, T. E. and Rubin, D. B. (1991). Large-sample significance levels from multiple imputed data using moment-based statistics and an F reference distribution. *Journal of the American Statistical Association*, **86**, 1065-1073.

Lipsitz, S. R. and Fitzmaurice, G. M. (1996). The score test for independence in $R \times C$ contingency tables with missing data. *Biometrics*, **52**, 751-762.

Lipsitz, S. R., Parzen, M. and Molenberghs, G. (1998). Obtaining the maximum likelihood estimates in incomplete $R \times C$ contingency tables using a Poisson generalized model. *Journal of Computational and Graphical Statistics*, **7**, 356-376.

Little, R. J. A. (1982). Models for nonresponse in sample surveys. *Journal of the American Statistical Association*, **77**, 237-250.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical analysis with missing data*, John Wiley & Sons, New York.

Magder, L. S. (2003). Simple approaches to assess the possible impact of missing outcome information on estimates of risk ratios, odds ratios and risk differences. *Controlled Clinical Trials*, **24**, 411-421.

Meng, X. L. and Rubin, D. B. (1991) Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical Association*, **86**, 899-909.

Meng, X. L. and Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, **79**, 103-111.

Meng, X. L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, **80**, 267-278.

Molenberghs, G., Kenward, M. G. and Goetghebeur, E. (2001). Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case. *Applied Statistics*, **50**, 15-29.

Pringle, R. M. and Rayner, A. A. (1971). *Generalized inverse matrices with applications to statistics*, Charles Griffin & Company Limited, London.

Rubin, D. B. (1976). Infernece and missing data (with discussion). *Biometrika*, **63**, 581-592.

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.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys,* John Wiley & Sons, New York.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, **91**, 473-489.

Rubin, D. B., Stern, H. S. and Vehovar, V. (1995). Handling don't know survey responses: The case of the slovenian plebiscite. *Journal of the American Statistical Association*, **90**, 822-828.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*, Chapman & Hall, London.

Shao, J. (1999). *Mathematical Statistics*, Springer-Verlag, New York.

Van Dyk, D. A., Meng, X. L. and Rubin, D. B. (1995). Maximum likelihood estimation via the ECM algorithm: Computing the asymptotic variance. *Statistica Sinica*, **5**, 55-75.

Wakefield, J. (2004). Ecological inference in 2x2 tables. *Journal of the Royal Statistical Society, Series A*, **167**, 385-426.

West, C. P. and Dawson, J. D. (2002). Complete imputation of missing repeated categorical data: One-sample applications. *Statistics in Medicine*, **21**, 203-217.