



Plain Answers to Several Questions about Association/Independence Structure in Complete/Incomplete Contingency Tables

S. K. Ghoreishi^{†,*} and M. R. Meshkani[‡]

[†] Qom University

[‡] Shahid Beheshti University

Abstract. In this paper, we develop some results based on Relational model (Klimova, et al. 2012) which permits a decomposition of logarithm of expected cell frequencies under a log-linear type model. These results imply plain answers to several questions in the context of analyzing of contingency tables. Moreover, determination of design matrix and hypothesis-induced matrix of the model will be discussed. Properties of maximum likelihood estimators of the model parameters are obtained. Some new model residuals and an alternative symmetric chi-square criterion are given. Two real examples illustrate the method.

Keywords. Association models; Context Specific models; Graphical models; Log-linear models; Relational models.

MSC 2010: 62H17, 62H12.

1 Introduction

Modeling expected cell frequencies in a given contingency table is a main topic in the literature. Various statistical models such as log-linear models

* Corresponding author

(Agresti, 2002) and association models (Goodman, 1972, 1985, 1991) provide, as far as possible, reasonable answers to several questions or hypotheses about global association/independence structure among categorical variables making up the contingency table. However, they seem to have some applicable limitations to test local association/independence across selected cells of the table. To deal with this problem, some alternative models: Graphical Models (Lauritzen, 1996), Hierarchical Subspace Models (Hara, et al. 2012), Relational Models (Klimova, et al. 2012), and Context Specific Interaction Models (Højsgaard, 2004) play an important role in this connection. Using these models, one can model expected cell frequencies for a given subset of cells which may have some characteristics in common. In spite of this prominent specialization, analysis approach via each model is apart from the others.

The main objective of this paper is to develop some results based on Relational models (Klimova, et al. 2012) which permits an orthogonal decomposition of logarithm of expected cell frequencies. They provide plain answers to complex questions about local association/independence structure as well as global one in any complete or incomplete contingency table. Indeed, this approach unifies all pictures in a unique one. Moreover, it produces a one to one relation between model and a contrast or generalized odds ratio under a given hypothesis.

Section 2 is devoted to basic definitions and notations. Since our approach is based on orthogonal decomposition of logarithm of expected cell frequencies, we need to determine its components, i.e., design matrix and its corresponding orthogonal matrix, called hypothesis-induced matrix. It is important to note that the design matrix depends on the parametrization used and, thus, cannot be determined uniquely. However, for many applications, having one of its form will be useful. Section 3 covers the subject in depth.

Section 4 considers the maximum likelihood estimation of model parameters and its properties under a given hypothesis. A new model residuals, alternative to Pearson residuals, and a symmetric chi-square distance, alternative to Pearson chi-square criterion, are introduced. Moreover, we discuss conditions under which, whenever they hold, one can estimate the expected cell frequencies via closed expressions.

Section 5 generalizes the results of Section 4 to general hypotheses. Finally, Section 6 illustrates the proposed method by analyzing two real data sets.

2 Definitions and Notations

In a multi-way contingency table, let X_1, \dots, X_k denote the categorical random variables with I_1, \dots, I_k levels and domains $\chi_1, \dots,$ and χ_k , respectively. Suppose $k \geq 1$ and $I_i \geq 2; i = 1, \dots, k$. The point $\mathbf{x} = (x_1, \dots, x_k) \in (\chi_1 \times \dots \times \chi_k)$ defines a cell in the contingency table, if and only if the event \mathbf{x} is not impossible in the population of interest. If \mathbf{x} logically cannot occur, it is a cell with *structural zero*.

Without loss of generality, suppose $|I|$ is the number of structurally non-zero cells of the table. Assume that the cells in the table are ordered lexicographically according to the levels of the categorical variables.

The i th cell parameter is denoted by τ_i . In a multinomial sampling design, $\tau_i = p_i$ while in a poisson sampling design $\tau_i = \lambda_i$.

The saturated model for cell probabilities under both designs can be written as (Klimova, et al. 2012)

$$\log \boldsymbol{\tau} = \mathbf{X}'\boldsymbol{\beta} + \mathbf{D}'\boldsymbol{\theta} \quad (1)$$

where $\log \boldsymbol{\tau} = (\log \tau_1, \dots, \log \tau_{|I|})$, here and in the sequel, is a vector-value function. Matrix \mathbf{X} is a matrix of full row rank with dimensions $J \times |I|$; $J < |I|$, having 0, 1 entries and is called design matrix. Matrix \mathbf{D} is a full row rank matrix with dimensions $(|I| - J) \times |I|$, whose entries are determined by the given null hypothesis.

Model (1) was motivated by traditional log-linear model, quasi models, topological models, indicator models, rater agreement-disagreement models, and two-way subtable sum models, see Klimova et al. (2012). This model covers almost all assumptions about cell's probabilities and association structures. Examples 1 and 2, in below, illustrate the subject with more details and give a deep insight on this model.

By assumptions, the matrices \mathbf{X} and \mathbf{D} are orthogonal to each other in the sense $\mathbf{D}\mathbf{X}' = \mathbf{0}$, $\mathbf{X}\mathbf{D}' = \mathbf{0}$. Thus, $\mathbf{M} = (\mathbf{X}', \mathbf{D}')'$ is a full rank matrix of rank $|I|$. The vectors $\boldsymbol{\tau}$, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$ are the parameters of distribution. It is important to note that under a given null hypothesis, both matrices \mathbf{X} and \mathbf{D} are known and the (1) is a saturated model. This representation is applicable to each complete or incomplete contingency table.

Let n_i be the i th cell observed frequency; $i = 1, \dots, |I|$, and $n = \sum_{i=1}^{|I|} n_i$. Under the saturated model (1) the maximum likelihood estimates of $\boldsymbol{\beta}$ and

θ are obtained from

$$\log \tau = \mathbf{X}'\beta + \mathbf{D}'\theta = \begin{pmatrix} \mathbf{X}' & \mathbf{D}' \end{pmatrix} \begin{pmatrix} \beta \\ \theta \end{pmatrix} = \mathbf{M}'\beta^*$$

as

$$\hat{\beta}^* = (\mathbf{M}\mathbf{M}')^{-1}\mathbf{M}\log \hat{\tau}, \quad (2)$$

with

$$\hat{\tau}_i = \begin{cases} \hat{p}_i = \frac{n_i}{n}, & \text{multinomial sampling} \\ n_i, & \text{poisson sampling.} \end{cases}$$

It is easily seen that the asymptotic distribution of $\hat{\beta}^*$ is a multivariate normal with mean β^* and some covariance matrix Σ . This distribution can be used for constructing confidence regions for β , θ and any linear combinations of theirs.

Here, it seems there are three points to note. First, if one wishes to test $H_0 : \theta = \theta_0$, perhaps a simple and natural way is to check whether θ_0 lies in the confidence region alluded to. But, is there a preferred way to employ the information provided by H_0 in obtaining β^* ?

Second, under H_0 the matrix \mathbf{D} is known. As it is evident from the following equation obtained from (1):

$$\mathbf{D}\log \tau = \mathbf{D}\mathbf{D}'\theta \iff \theta = (\mathbf{D}\mathbf{D}')^{-1}\mathbf{D}\log \tau. \quad (3)$$

In spite of this fact, is there any way to find matrix \mathbf{X} corresponding to H_0 , in order to carry over the information provided by H_0 into \mathbf{X} ?

Substituting \mathbf{p} for τ , it is seen that (3) is in fact a contrast or a generalized odds ratio. Thus, the null hypothesis H_0 states that this contrast holds.

The third point regards the question of the order of estimation for parameters and expected cell frequencies. Under given null hypothesis, should one estimate the model parameters first and then obtain the expected cell frequencies or there is a way to estimate expected cell frequencies directly in closed form?

In the rest of the paper, we address these three points.

3 Determination of Matrices \mathbf{X} and \mathbf{D}

In this section, we determine \mathbf{X} and \mathbf{D} based on a given hypothesis. First, some simple illustrative examples.

Example 1. Consider a 2×2 table with cell probabilities $\mathbf{p} = (p_{11}, p_{12}, p_{21}, p_{22})' \equiv (p_1, p_2, p_3, p_4)'$. We wish to test the hypothesis $H_{01} : p_4 = \frac{1}{5}$ or equivalently, $H_{01} : \log p_4 = -\log 5$. Thus \mathbf{X} and \mathbf{D} , as explained later, are

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{D} = (0, 0, 0, 1).$$

Next, suppose, $H_{02} : \frac{p_2}{p_3} = 1$ or $H_{02} : \log p_2 - \log p_3 = 0$ is to be tested. Hence, as we shall see later,

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{D} = (0, 1, -1, 0).$$

In this example, the equivalent hypotheses in terms of $\boldsymbol{\theta}$ in model (1) are $H_{01} : \boldsymbol{\theta} = -\log 5$ and $H_{02} : \boldsymbol{\theta} = 0$, respectively.

Example 2. Consider a $2 \times 2 \times 2$ contingency table and an independence model of $[Y_1 Y_2]$ and $[Y_3]$. The probability model for this table is $p_{ijk} = p_{ij+} \cdot p_{++k}$, where $+$ denotes summed over the corresponding index. Put

$$\mathbf{p} = (p_{111}, p_{112}, p_{121}, p_{122}, p_{211}, p_{212}, p_{221}, p_{222})' \equiv (p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8)'$$

Then, as we shall see later,

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad a \ 5 \times 8 \text{ matrix,}$$

and

$$\mathbf{D} = \begin{pmatrix} 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 \end{pmatrix}, \quad a \ 3 \times 8 \text{ matrix.}$$

Here, the hypothesis H_0 in the model (1) is equivalent to $H_0 : \boldsymbol{\theta} = (0, 0, 0)'$.

In the above examples, the design matrix \mathbf{X} and hypothesis-induced matrix \mathbf{D} were determined by some efforts. In general, we need some rules

to determine them. These rules will be discussed later. To this end, we need some definitions.

Imposition of some constraints on the logarithms of the cell probabilities can be expressed as

$$H_0 : \sum_{i=1}^{|I|} d_i(j) \log p_i = \delta_j; \quad j = 1, \dots, |I| - J, \quad (4)$$

where the coefficients $d_i(j)$ and quantities δ_j are assumed to be known.

Definition 1. If $d_i(j)$ satisfy in

$$\sum_{i=1}^{|I|} d_i(j) = 0; \quad j = 1, \dots, |I| - J,$$

then (4) is called a contrast or generalized odds ratio. Let

$$\mathbf{d}(j) = (d_1(j), \dots, d_{|I|}(j))'; \quad \text{for } j = 1, \dots, |I| - J.$$

Remember that the number of independent contrast under H_0 is equal to the number of degrees of freedom for the χ^2 statistic. Also note that $\mathbf{d}(j)$, indeed, constitutes the rows of the hypothesis-induced matrix \mathbf{D} . Thus, from (1) and (4), it is obvious that δ_j is the j th element of vector $\mathbf{D}\mathbf{D}'\boldsymbol{\theta}$ and an alternative form for the hypothesis to (4) is

$$H_0 : \boldsymbol{\theta} = (\mathbf{D}\mathbf{D}')^{-1}\boldsymbol{\delta},$$

where $\boldsymbol{\delta} = (\delta_1, \dots, \delta_{|I|-J})'$.

Maybe, it seems that our biggest challenge, in H_0 , is the determination of $\boldsymbol{\delta}$. But, essentially, it is not a serious problem. Because, the analyst's subject ideas, that are given by (4), automatically, imply the quantity $\boldsymbol{\delta}$. Examples 1-3 support this claim.

A contrast is said to be of *simple first type* whenever elements of $\mathbf{d}(j)$ for a specific j are all zero except two of them which are 1 and -1. If there are four non-zero elements in $\mathbf{d}(j)$ (two +1 and two -1), then it is a *simple second type* contrast. Other types are called *third type* contrast. Note that $\mathbf{D}\mathbf{1} = \mathbf{0}$, where $\mathbf{1} = (1, \dots, 1)'$, and since $\mathbf{X}\mathbf{D}' = \mathbf{0}$, it follows that vector $\mathbf{1}$ lies in the row space of design matrix \mathbf{X} . Here, we concentrate on the contrasts of first and second type.

Assuming $\mathbf{d}(j)$ is a first or a second type contrast, without loss of generality, we would respectively have $\mathbf{d}(j) = (1, -1, 0, \dots, 0)'$ or $\mathbf{d}(j) = (1, -1, -1, 1, 0, \dots, 0)'$.

Definition 2. The two contrasts $\sum_{i=1}^{|I|} d_i(j) \log p_i = 0$ and $\sum_{i=1}^{|I|} d_i(j') \log p_i = 0$ are said to be independent if their coefficients are orthogonal to each other, i.e., $\sum_{i=1}^{|I|} d_i(j)d_i(j') = 0$; for $j \neq j'$.

It is important to note that the independent contrasts are come out by assuming some linear structures on completely separated subsets of cell's probabilities. Moreover, these contrasts decompose the global Chi-square statistics into some additive terms, Goodman (1985).

If contrasts are independent under H_0 , one can immediately conclude that the cells forming different contrasts constitute separate sets. For example, in a 3×3 contingency table, the two second type contrasts $\log p_{12} - \log p_{22} - \log p_{13} + \log p_{23} = 0$ and $\log p_{21} - \log p_{31} - \log p_{22} + \log p_{32} = 0$ share the cell (2, 2), thus they are not independent.

Below, we shall discuss how to determine the design matrix \mathbf{X} under H_0 for both dependent and independent contrasts. The four cases will be treated, separately. Note that each contrast is identified by the vector of its coefficients, $\mathbf{d}(j)$.

3.1 The First Type Contrasts

Consider two specific first type contrast, $\mathbf{d}(j)$ and $\mathbf{d}(j')$. They may be either independent or share at most one cell probability. The following algorithm may be used for constructing a design matrix:

Initially, take an identity matrix \mathbf{U} of order $|I| \times |I|$. Number its rows and columns according to probability vector \mathbf{p} in the model (1). Perform the following summing operations on its rows.

- Case 1. If contrast $\mathbf{d}(j)$ and $\mathbf{d}(j')$, $j \neq j'$ are independent, form a substitute row by adding those two rows in \mathbf{U} whose numbers correspond to non-zero entries of $\mathbf{d}(j)$. Do the same operation for $\mathbf{d}(j')$. Each operation would reduce the number of rows in \mathbf{U} by one, hence yielding a new matrix, \mathbf{X} , of order $(|I| - 2) \times |I|$.
- Case 2. If the contrasts $\mathbf{d}(j)$ and $\mathbf{d}(j')$, $j \neq j'$ are dependent, they surely are made of three cell probabilities of the contingency table.

In this case, three rows in \mathbf{U} whose numbers correspond to non-zero entries of $\mathbf{d}(j)$ and $\mathbf{d}(j')$, $j \neq j'$ are added to form one row. Thus \mathbf{X} is found with order $(|I| - 2) \times |I|$.

The number of omitted rows of \mathbf{U} is equal to the degrees of freedom of chi-square statistic. Having performed the above operations for all contrasts under H_0 , it would lead to a design matrix, \mathbf{X} , for H_0 with J rows whose row space contains vector $\mathbf{1}$.

3.2 The Second Type Contrast

Consider two specific contrasts $\mathbf{d}(j)$ and $\mathbf{d}(j')$, $j \neq j'$, of the second type. They may be either independent or share at most two cells of the table. The design matrix \mathbf{X} can be constructed by the following algorithm: Initially, take an identity matrix \mathbf{U} of order $|I| \times |I|$. Number its rows and columns as described in Section 3.1.

- Case 1. If the two contrasts are independent, for each contrast, choose four rows in \mathbf{U} whose numbers correspond to the row indices of non-zero entries of $\mathbf{d}(j)$. Replace these four rows by the following three rows:
 1. The sum of all those four chosen rows in \mathbf{U} .
 2. The sum of the first and the second chosen rows in \mathbf{U} .
 3. The sum of the first and the third chosen rows in \mathbf{U} .

Do the same for the second contrast. Clearly, each of these operations reduces the number of rows in \mathbf{U} by one. Hence, we obtain a matrix of order $(|I| - 2) \times |I|$

- Case 2. In case the contrasts are dependent and they share one cell probability, they are made of 7 distinct cell probability altogether. To account for these 2 contrasts in the chi-square statistic, one should reduce its degrees of freedom by 2. Without loss of generality, let the coefficient of the common term in both contrasts be +1. Consider the 7 rows in \mathbf{U} corresponding to row indices present in the contrasts. Make 5 substitute rows from those 7 by the following operations:
 1. The sum of all those 7 rows in \mathbf{U} .

2. The sum of the rows in \mathbf{U} whose number is the same as the row index of the common cell probability and the other two rows with coefficients -1 in each contrast.
 3. Similar to Step 1 but use two rows in \mathbf{U} with coefficients -1 in each contrast which have not already been used.
 4. The sum of two rows in \mathbf{U} whose numbers appear in the first contrast, their coefficients have opposite signs, and neither contain the common cell probability.
 5. Similar to Step 4 but use the two rows in \mathbf{U} whose numbers appear in the second contrast.
- Case 3. If the two contrasts share two cell probabilities, they are made of 6 distinct cell probabilities. These two contrasts reduce the degrees of freedom of chi-square statistic by two. Hence, one should replace those 6 rows in \mathbf{U} which make up the two contrasts by 4 rows, as follows

1. Suppose the common cell probabilities present in the contrasts have opposite signs. Without loss of generality, suppose the second and fifth cell probabilities are common. Suppose the contrasts expressed in logarithms of adjacent odds ratios constitute the 2×3 table

$$\mathbf{X}_0 = \begin{pmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \end{pmatrix}.$$

Then the substitute rows are formed as

- (a) Put 1 for all entries of the columns in \mathbf{U} which correspond to cell probabilities 1 to 6 and 0 for the rest.
 - (b) Put 1 for all entries entries of the columns in \mathbf{U} which correspond to cell probabilities in the first column of \mathbf{X}_0 and 0 for the rest.
 - (c) Put 1 for all entries of the columns in \mathbf{U} which correspond to cell probabilities in the second column of \mathbf{X}_0 and 0 for the rest.
2. Suppose common cell probabilities present in the contrasts have the same sign. In this case, \mathbf{X} can not be constituted since then \mathbf{X} and \mathbf{D} will not be orthogonal to each other.

The following examples illustrate the above algorithm.

Example 3. Referring to Example 1, suppose we wish to test $H_0 : p_1 = p_2 = p_3$ or equivalently the contrasts $H_0 : \log p_1 = \log p_2$ and $\log p_1 = \log p_3$. These dependent first type contrasts share $\log p_1$. Following the algorithm given in case 2, we get

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \end{pmatrix}.$$

Example 4. Consider the following 3×3 incomplete table

$$\begin{pmatrix} - & p_1 & p_2 \\ p_3 & p_4 & p_5 \\ p_6 & p_7 & - \end{pmatrix}.$$

We wish to test $H_0 : \frac{p_1 p_5}{p_2 p_4} = 1$ and $\frac{p_3 p_7}{p_6 p_4} = 1$ or what is the same

$$H_0 : \log p_1 + \log p_5 - \log p_2 - \log p_4 = 0, \quad \log p_3 + \log p_7 - \log p_6 - \log p_4 = 0$$

that is, $H_0 : \boldsymbol{\theta} = (0, 0)'$.

These two contrasts are dependent and of type 2 which share p_4 of cell (2, 2). By the algorithm given in case 2, we get

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} 1 & -1 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{pmatrix}.$$

Example 5. Consider the 2×3 table

$$\begin{pmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \end{pmatrix}.$$

Following contrasts are an example of the case where they are dependent and share 2 cell probabilities p_2 and p_5 .

$$H_0 : \log p_1 + \log p_5 - \log p_2 - \log p_4 = 0, \quad \log p_2 + \log p_6 - \log p_5 - \log p_3 = 0$$

that is, $H_0 : \boldsymbol{\theta} = (0, 0)'$.

By the algorithm given in case 2, we have

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} 1 & -1 & 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 & -1 & 1 \end{pmatrix}.$$

4 Maximum Likelihood Estimates of $\boldsymbol{\beta}$ under $H_0 : \boldsymbol{\theta} = \mathbf{0}$

Under the hypothesis $H_0 : \boldsymbol{\theta} = \mathbf{0}$, the model (1) reduces to

$$\log \boldsymbol{\tau} = \mathbf{X}'\boldsymbol{\beta}. \quad (5)$$

Model (5) is equivalent to a Relational model where in the context of log-linear models, the design matrix \mathbf{X} is obtained from the cylinders, corresponding marginal totals, given by problem of interest. Using the notation given in Klimova et al. (2012), the model (5) can be represented as $RM(p)$ for $\tau_i = p_i$ and as $RM(\lambda)$ for $\tau_i = \lambda_i$. As we saw in Section 3, when H_0 involves contrasts which satisfy (4), then vector $\mathbf{1}$ is in the row space of \mathbf{X} . This property entails $RM(p)$ to have a regular exponential family representation with dimension $J - 1$ and $RM(\lambda)$ with dimension J , see Theorems 3.1 and 3.2 in Klimova et al. (2012).

Under both $RM(p)$ and $RM(\lambda)$ models the vector $\mathbf{X}\mathbf{n}$ is the sufficient statistic where $\mathbf{n} = (n_1, \dots, n_{|I|})$ is vector of observed frequencies in the table. When all entries of $\mathbf{X}\mathbf{n}$ are positive, MLE of $\boldsymbol{\beta}$ exist and it is unique. Moreover, MLE 's of cell expected frequencies, under these two models, are the same, see Theorems 4.1 and 4.3 in Klimova, et al. (2012). Therefore, without loss of generality, in the sequel, we shall use \mathbf{p} instead of $\boldsymbol{\tau}$. The likelihood equation for estimating the parameters of $RM(p)$ model is

$$\mathbf{X}\mathbf{n} = n\mathbf{X}\hat{\mathbf{p}} \quad (6)$$

where $\hat{\mathbf{p}}$ is the estimate of cell probabilities under the model (5).

From (6), we have

$$\mathbf{X}\mathbf{n} = n\mathbf{X}\hat{\mathbf{p}} \Leftrightarrow \mathbf{X}(\mathbf{n} - n\hat{\mathbf{p}}) = \mathbf{0}. \quad (7)$$

This result shows that $\mathbf{e} = \mathbf{n} - n\hat{\mathbf{p}}$, the vector of residuals, is a vector in the row space of matrix \mathbf{D} . That is, there exists a vector \mathbf{k} such that

$$\mathbf{e} = \mathbf{D}'\mathbf{k}.$$

Note that, in practice, Pearson residuals $R_i = \frac{n_i - n\hat{p}_i}{\sqrt{n\hat{p}_i}}$ are often used for residual analysis. From (7) it is evident that the model residuals satisfy $J - 1$ linear combinations. Thus, any residual analysis based on \mathbf{e} suffer from those dependencies. Therefore, we chose to use alternative residuals of the form $u_i = \log \frac{r_i}{\hat{p}_i}$; $i = 1, \dots, |I|$ where $r_i = \frac{n_i}{n}$. Properties of u_i will be discussed in theorem 5. Here, we have some theorems.

It is obvious the model (5) is an especial kind of log-linear models. Thus, all properties about log-linear models satisfy here too. Among all, for a fixed number of cells, as $n \rightarrow \infty$, the ML estimator $\hat{\beta}$ is asymptotically normal with mean β and covariance matrix $\text{cov}(\hat{\beta})$. In Poisson sampling, we have $\text{cov}(\hat{\beta}) = \{\mathbf{X}'\text{diag}(\boldsymbol{\tau})\mathbf{X}\}^{-1}$, while in multinomial sampling scheme $\text{cov}(\hat{\beta}) = \{\mathbf{X}'(\text{diag}(\boldsymbol{\tau}) - \frac{\boldsymbol{\tau}\boldsymbol{\tau}'}{n})\mathbf{X}\}^{-1}$. For more details see Agresti (2002) Section 8.7.

Moreover, Algorithm (a), in Section 5, is applicable to obtain the ML estimate of $\hat{\beta}$ by assuming $\boldsymbol{\theta}_0 = \mathbf{0}$ and using (5).

Theorem 1. *Under the model (5), one has*

$$\mathbf{e}'\log \hat{\mathbf{p}} = \sum_{i=1}^{|I|} (n_i - n\hat{p}_i) \log \hat{p}_i = 0,$$

and consequently $\sum_{i=1}^{|I|} n_i \log \hat{p}_i = \sum_{i=1}^{|I|} n\hat{p}_i \log \hat{p}_i$.

Proof. From (5), $\log \hat{\mathbf{p}} = \mathbf{X}'\hat{\beta}$, so it is in the row space of \mathbf{X} . From (7), $\mathbf{e} = \mathbf{n} - n\hat{\mathbf{p}}$ is one of the vectors in the row space of \mathbf{D} . But \mathbf{X} and \mathbf{D} are orthogonal to each other, so are \mathbf{e} and $\log \hat{\mathbf{p}}$. \square

Theorem 2. *Under the model (5), one has*

$$\chi^2 = \sum_{i=1}^{|I|} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} \simeq \sum_{i=1}^{|I|} (n_i - n\hat{p}_i) \log r_i = \mathbf{e}'\log \mathbf{r} = \mathbf{e}'\log \mathbf{d}$$

where $\mathbf{r} = (\frac{n_1}{n}, \dots, \frac{n_{|I|}}{n})'$ and $\mathbf{d} = (\frac{r_1}{\hat{p}_1}, \dots, \frac{r_{|I|}}{\hat{p}_{|I|}})' = (\frac{n_1}{n\hat{p}_1}, \dots, \frac{n_{|I|}}{n\hat{p}_{|I|}})'$, the ratio of observed frequencies to the expected ones under the model (5).

Proof. Taylor series expansion of $\log r_i$ about the point \hat{p}_i is

$$\log r_i \simeq \log \hat{p}_i + \frac{1}{\hat{p}_i}(r_i - \hat{p}_i) = \log \hat{p}_i + \frac{1}{n\hat{p}_i}(n_i - n\hat{p}_i). \quad (8)$$

Substituting for $\frac{(n_i - n\hat{p}_i)}{n\hat{p}_i}$ from (8) in the χ^2 formula, we get

$$\begin{aligned} \chi^2 &= \sum_{i=1}^{|I|} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} = \sum_{i=1}^{|I|} (n_i - n\hat{p}_i) \frac{(n_i - n\hat{p}_i)}{n\hat{p}_i} \\ &\simeq \sum_{i=1}^{|I|} (n_i - n\hat{p}_i)(\log r_i - \log \hat{p}_i) \\ &= \sum_{i=1}^{|I|} (n_i - n\hat{p}_i) \log \frac{r_i}{\hat{p}_i} \\ &= \sum_{i=1}^{|I|} (n_i - n\hat{p}_i) \log d_i. \end{aligned}$$

□

We denote the above approximation by κ^2 . It is important to note that the above approximation is symmetric with respect to permutation of n_i and $n\hat{p}_i$. Also its positivity stems from the definition and positivity of the likelihood chi-square statistics and concavity of logarithm function.

Theorem 3. If $\hat{\beta}_F$ is the estimate of the model parameters under the full model (1) and $\hat{\beta}_R$ is those under the reduced model (5), Then

$$\hat{\beta}_F - \hat{\beta}_R = (\mathbf{X}\mathbf{X}')^{-1}\mathbf{X} \log \mathbf{d}.$$

Proof. For the saturated model,

$$\log \mathbf{r} = \mathbf{X}'\hat{\beta}_F + \mathbf{D}'\hat{\theta}_F.$$

But, under the reduced model (5), MLE $\hat{\beta}$ satisfies $\log \hat{\mathbf{p}} = \mathbf{X}'\hat{\beta}_R$. Thus,

$$\log \mathbf{r} - \log \hat{\mathbf{p}} = \log \frac{\mathbf{r}}{\hat{\mathbf{p}}} = \mathbf{X}'(\hat{\beta}_F - \hat{\beta}_R) + \mathbf{D}'\hat{\theta}_F.$$

Multiply both sides by \mathbf{D} and \mathbf{X} and use the orthogonality of them to get

$$\hat{\theta}_F = (\mathbf{D}\mathbf{D}')^{-1}\mathbf{D} \log \mathbf{d},$$

and finally,

$$\hat{\beta}_F - \hat{\beta}_R = (\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}\log\mathbf{d}.$$

□

Once again, consider the likelihood equation (6). Let $\mathbf{X}(j)$ denote the j th row of \mathbf{X} . Then (6) is

$$\mathbf{X}(j)\mathbf{n} = n\mathbf{X}(j)\hat{\mathbf{p}}; \quad j = 1, \dots, J. \quad (9)$$

Based on (9), we have the following theorem.

Theorem 4. *If under H_0 (either the first or the second type contrast defined in (4)) one has*

$$p_i = \frac{[\mathbf{X}(i_2)\mathbf{p}][\mathbf{X}(i_3)\mathbf{p}]}{\mathbf{X}(i_1)\mathbf{p}}; \quad i_1 \neq i_2 \neq i_3, \quad (10)$$

then the MLE for the i th cell expected frequency, $n\hat{p}_i$, has the closed form

$$n\hat{p}_i = \frac{[\mathbf{X}(i_2)\mathbf{n}][\mathbf{X}(i_3)\mathbf{n}]}{\mathbf{X}(i_1)\mathbf{n}}.$$

Moreover, (10) can be generalized to

$$p_i = \frac{\prod_{j=1}^u [\mathbf{X}(i_j)\mathbf{p}]}{\prod_{j'=1}^v [\mathbf{X}(i_{j'})\mathbf{p}]}; \quad v < u \text{ and } j' \neq j.$$

All closed formulas for various independence hypotheses (whether joint or conditional in the log-linear models) satisfy this general rule.

Proof. The proof immediately follows from application of (9). □

Relation (10), in Theorem 4, is popular in the literature, Hara et al. (2012). To see this, revisit Example 2. Assuming three row vectors $\mathbf{X}(1_1) = (1, 1, 1, 1, 1, 1, 1, 1)$, $\mathbf{X}(1_2) = (1, 1, 0, 0, 0, 0, 0, 0)$, and $\mathbf{X}(1_3) = (1, 0, 1, 0, 1, 0, 1, 0)$, it is obvious that $p_{11+} = \mathbf{X}(1_2)\mathbf{p}$, $p_{++1} = \mathbf{X}(1_3)\mathbf{p}$, and $1 = \mathbf{X}(1_1)\mathbf{p}$. So, using corresponding notations in Example 2, we have

$$p_{111} = p_{11+}p_{++1} \Leftrightarrow p_1 = \frac{[\mathbf{X}(1_2)\mathbf{p}][\mathbf{X}(1_3)\mathbf{p}]}{\mathbf{X}(1_1)\mathbf{p}}.$$

This means that (10) satisfies for joint independence in a $2 \times 2 \times 2$ contingency table.

Theorem 5. According to definition of \mathbf{r} and $\hat{\mathbf{p}}$, the vector $\log \mathbf{d} = (\log \frac{r_1}{\hat{p}_1}, \dots, \log \frac{r_{|I|}}{\hat{p}_{|I|}})'$ has a asymptotic normal distribution with mean $\mathbf{0}$ and and covariance matrix $\Sigma = [\sigma_{ij}]$ where

$$\sigma_{ij} = \begin{cases} \frac{1-p_i}{np_i} & i = j \\ -\frac{1}{n} & i \neq j. \end{cases}$$

Proof. Let \hat{p}_i be replaced by p_i in (8). i.e.,

$$\log r_i \simeq \log p_i + \frac{1}{p_i}(r_i - p_i) = \log p_i + \frac{1}{np_i}(n_i - np_i).$$

Then

$$\log \frac{r_i}{p_i} \simeq \frac{1}{p_i}(r_i - p_i).$$

Multiplying each side by n and p_i yields

$$np_i \log \frac{r_i}{p_i} \simeq (n_i - np_i).$$

But, we know $(n_i - np_i)$ are asymptotically distributed as normal with mean 0 and variance $np_i(1 - p_i)$. Also

$$\text{cov}(n_i - np_i, n_j - np_j) = -np_i p_j, \quad i \neq j.$$

Hence

$$\log \frac{r_i}{p_i} \sim N\left(0, \frac{1-p_i}{np_i}\right),$$

and

$$\text{cov}\left(\log \frac{r_i}{p_i}, \log \frac{r_j}{p_j}\right) = \frac{-1}{n}, \quad i \neq j.$$

By replacing \hat{p}_i for p_i and resorting to Cramer-Wold device (Serfling, 1980), proof is completed. \square

Regarding Theorem 5, one can claim that as $n \rightarrow \infty$, the correlation between the residuals $\log \frac{r_i}{\hat{p}_i}$ tend to zero. Furthermore, the standardized residuals will be distributed as

$$\text{Std. } u_i = \sqrt{\frac{n\hat{p}_i}{1 - \hat{p}_i}} \log \frac{r_i}{\hat{p}_i} \sim N(0, 1).$$

Our numerical study, unpolished, shows that these residuals tend to have greater variance with respect to the Pearson residuals and, moreover, Pearson residuals converge more slowly than our residuals to normal. As we know Pearson residuals are asymptotically normal with mean 0 but their asymptotic variance is less than 1.0. Indeed, comparing Pearson residuals to standard normal percentage points provides conservative indications of cells having lack of fit, Agresti (2002). It seems our residuals to be a good alternative to Pearson residuals. Here, we give up from more discussion about their details and leave it as an open problem.

5 Maximum Likelihood Estimates under the General Hypothesis $\boldsymbol{\theta} = \boldsymbol{\theta}_0$

When the vector $\mathbf{1}$ belongs to the row space of \mathbf{X} , the expected cell frequencies under both multinomial and poisson sampling will be the same. Multiplying (1) through by \mathbf{D} to get

$$\mathbf{D} \log \mathbf{p} = \mathbf{D}\mathbf{D}'\boldsymbol{\theta} \Leftrightarrow \boldsymbol{\theta} = (\mathbf{D}\mathbf{D}')^{-1}\mathbf{D} \log \mathbf{p}. \quad (11)$$

Thus, $\boldsymbol{\theta}$ is in fact a generalized odds ratio. Consider testing: $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$. This hypothesis states that $J - |I|$ generalized odds ratios are equal to arbitrary values, specified by the hypothesis of interest, not all of them zero. Under H_0 , MLE of \mathbf{p} is given below.

Theorem 6. For the models (1) and under H_0 , MLE of \mathbf{p} satisfies the equations:

$$\begin{cases} \mathbf{X}\mathbf{n} = n\mathbf{X}\hat{\mathbf{p}}, \\ (\mathbf{D}\mathbf{D}')^{-1}\mathbf{D} \log \hat{\mathbf{p}} = \boldsymbol{\theta}_0. \end{cases} \quad (12)$$

Proof. In the model (1) and under H_0 , the cell probabilities satisfy the equations

$$\log p_i = \mathbf{X}'(i)\boldsymbol{\beta} + \mathbf{D}'(i)\boldsymbol{\theta}_0; \quad i = 1, \dots, |I|$$

where $\mathbf{X}'(i)$ and $\mathbf{D}'(i)$ are the i th column of \mathbf{X} and \mathbf{D} , respectively. Substitute these values in the likelihood function and ignore the constants related

to the constraint $\sum_{i=1}^{|I|} p_i = 1$. Then

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{|I|} n_i (\mathbf{X}'(i)\boldsymbol{\beta} + \mathbf{D}'(i)\boldsymbol{\theta}_0) - \eta \left(\sum_{i=1}^{|I|} \{\exp(\mathbf{X}'(i)\boldsymbol{\beta} + \mathbf{D}'(i)\boldsymbol{\theta}_0) - 1\} \right)$$

which is to be maximized subject to $\mathbf{1}'\mathbf{p} = 1$. Thus, we arrive at

$$\begin{cases} \mathbf{X}\mathbf{n} = \eta\mathbf{X}\hat{\mathbf{p}} \\ \mathbf{1}'\mathbf{p} = 1, \end{cases}$$

where η is the Lagrange multiplier. Since vector $\mathbf{1}$ is in the row space of \mathbf{X} , there exists a vector \mathbf{k} such that $\mathbf{k}'\mathbf{X} = \mathbf{1}'$. Hence, by multiplying the first equation through by \mathbf{k}' , it is obtained $\eta = n$. The equation (12) is the direct result of (11). \square

Although Theorem 6 provides an explicit solution for \mathbf{p} or equivalently for the expected cell frequencies $n\hat{\mathbf{p}}$, one may encounter some complications in practice. So, we provide the following algorithm.

Algorithm (a):

- Initially, obtain the preliminary estimate of $\boldsymbol{\beta}$ from $\tilde{\boldsymbol{\beta}} = (\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}\log\mathbf{r}$.
- Obtain the preliminary estimate of \mathbf{p} from $\log\tilde{\mathbf{p}} = \mathbf{X}'\tilde{\boldsymbol{\beta}} + \mathbf{D}'\boldsymbol{\theta}_0$, for given $\boldsymbol{\theta}_0$.
- Use the initial value $\mathbf{p}^{(0)} = \tilde{\mathbf{p}}$ to iterate,

$$p_i^{(t+j)} = p_i^{(t+j-1)} \frac{\mathbf{X}(j)\mathbf{n}}{n\mathbf{X}(j)\mathbf{p}^{(t+j-1)}}; \quad i = 1, \dots, |I|, \quad j = 1, \dots, J, \quad t = 0, 2, 3, \dots$$

- Iterate for enough large times to reach convergence. According to the convergence property of iterative proportional method, see Agresti (2002), as $N \rightarrow \infty$,

$$p_i^{(N)} \rightarrow \hat{p}_i; \quad i = 1, 2, \dots, |I|,$$

and \hat{p}_i 's satisfy (12).

In case, one drops the assumption that vector $\mathbf{1}$ belongs to the row space of \mathbf{X} , the equations (12) would become

$$\begin{cases} \mathbf{X}\mathbf{n} = \eta\mathbf{X}\hat{\mathbf{p}}, \\ (\mathbf{D}\mathbf{D}')^{-1}\mathbf{D}\log\hat{\mathbf{p}} = \boldsymbol{\theta}_0, \\ \mathbf{1}'\mathbf{p} = 1. \end{cases} \quad (13)$$

In this case, \mathbf{D} is not orthogonal to vector $\mathbf{1}$ and the contrasts are not of the first and second type under H_0 . The following algorithm can be used for this case to obtain *MLE* of expected cell frequencies and η .

Algorithm (b):

- Set $m_i = \eta p_i$; $i = 1, \dots, |I|$
- Choose arbitrary initial values $m_i^{(0)} > 0$ such that they satisfy the second equation in (13).
- From the first equation of (13), for j th row of \mathbf{X} we have

$$\mathbf{X}(j)\mathbf{n} = \mathbf{X}(j)\hat{\mathbf{m}},$$

so use given $m_i^{(0)}$ and update them for $i = 1, \dots, |I| - J$, from

$$m_i^{(t+j)} = m_i^{(t+j-1)} \frac{\mathbf{X}(j)\mathbf{n}}{\mathbf{X}(j)\mathbf{m}^{(t+j-1)}}; \quad j = 1, \dots, J, \quad t = 0, 2, 3, \dots$$

In each stage, find $m_i^{(t+j)}$, ($i = |I| - J, \dots, |I|$) such that $(\mathbf{D}\mathbf{D}')^{-1}\mathbf{D}\log\hat{\mathbf{p}} = \boldsymbol{\theta}_0$

- According to the convergence property of iterative proportional method, as $N \rightarrow \infty$, $m_i^{(N)} \rightarrow \hat{m}_i$ and consequently, one has

$$\hat{\eta} = \frac{\sum_{i=1}^{|I|} n_i}{\sum_{i=1}^{|I|} \hat{m}_i},$$

which provides the expected cell frequencies under H_0 as $\hat{\eta}\hat{m}_i$.

6 Application

We illustrate the above results by analyzing two real data sets. These data have also been analyzed by some other statisticians. But our analysis is based on the proposed method, specially the last two iterative algorithms given in Section 5.

Example 6. Table 1 shows the changes in responds' job with respect to their fathers'. Suppose, we are interested in testing the following equivalent hypotheses: $H_0 : \frac{p_{12}p_{23}}{p_{13}p_{22}} = 0.1, \frac{p_{21}p_{32}}{p_{31}p_{22}} = 0.1,$

$$H_0 : \begin{cases} \log p_{12} + \log p_{23} - \log p_{13} - \log p_{22} = -2.30 \\ \log p_{21} + \log p_{32} - \log p_{31} - \log p_{22} = -2.30, \end{cases}$$

or $H_0 : \theta = (-0.6, -0.3)'$. Matrices \mathbf{X} and \mathbf{D} are given in Example 4.

Table 1. Occupation statuses of respondents and their fathers, Klimova et al. (2012)

Father's Job	Respondent's occupation		
	White-collar	Manual	Farm
White-collar	–	6313	2776
Manual	6321	10883	294
Farm	8619	2471	–

Using Algorithm (a), we have computed the expected cell frequencies under H_0 , Pearson residuals, u -residuals, and their standardized values. They are shown in Table 2 along with the observed cell frequencies.

Table 2. Estimates of expected cell frequencies and their residuals

Pearson residuals	$Std.u_i = \sqrt{\frac{n\hat{p}_i}{1-\hat{p}_i}} \log \frac{n_i}{n\hat{p}_i}$	$u_i = \log \frac{n_i}{n\hat{p}_i}$	$n\hat{p}_i$	n_i
-2.11	-2.36	-0.026	6482.77	6313
1.18	1.21	0.022	2714.72	2776
0.48	0.51	0.0058	6284.24	6321
1.50	1.76	0.0144	10727.60	10883
-1.45	-1.46	-0.026	301.91	294
0.93	1.07	-0.010	8706.03	8619
0.29	0.30	0.006	2456.63	2471

Under hypothesis H_0 , Pearson χ^2 statistic is 9.456 while our approximate value is $\kappa^2 = 9.49$ with p-value 0.008. The standard residuals show that the lack of fit is due to cell (1,2) with large standard residual -2.36.

Pearson χ^2 for the hypothesis

$$H_0 : \frac{p_{12}p_{23}}{p_{13}p_{22}} = 1, \quad \frac{p_{21}p_{32}}{p_{31}p_{22}} = 1,$$

is equal to $\chi^2 = 6995.83$. For more details of analyzing of these data see Klimova, et al. (2012).

Example 7. (Agresti, 2002, P.25). Table 3 provides the data on 156 dairy calves classified according to whether they caught pneumonia within 60 days of birth. Calves that got a pneumonia infection were also classified according to whether they got a secondary infection within two weeks after the first infection cleared up.

Although the researchers were interested in testing the hypothesis $p_{11} + p_{12} = \frac{p_{11}}{p_{11}+p_{12}}$ which reduces to $\frac{p_{11}p_{22}^2}{p_{12}^2} = 1$, we consider a general hypothesis $H_0 : \frac{p_{11}p_{22}^2}{p_{12}^2} = 0.2$. Its equivalent form is

$$H_0 : \log p_{11} + 2 \log p_{22} - 2 \log p_{12} = -1.61,$$

or $H_0 : \boldsymbol{\theta} = -0.18$. Regarding this hypothesis,

$$\mathbf{X} = \begin{pmatrix} 0 & 1 & 1 \\ 2 & 1 & 0 \end{pmatrix}, \quad \mathbf{D} = (1, \quad -2, \quad 2).$$

Using Algorithm (b), the results are shown in Table 4. Pearson $\chi^2 = 0.0119$ and $\kappa^2 = 0.0014$ with p-value 0.91 and $\hat{\eta} = 0.998$. Thus H_0 cannot be rejected. Table 4 shows the details.

Table 3. Primary and secondary pneumonia infection in calves

Primary Infection	Secondary Infection	
	Yes	No
Yes	30	63
No	–	63

Table 4. Estimates of expected cell frequencies and their residuals

Pearson residuals	$Std.u_i = \sqrt{\frac{n\hat{p}_i}{1-\hat{p}_i}} \log \frac{n_i}{n\hat{p}_i}$	$u_i = \log \frac{n_i}{n\hat{p}_i}$	$n\hat{p}_i$	n_i
-0.041	-0.046	-0.007	30.23	30
0.0847	0.109	0.011	62.33	63
-0.055	-0.072	-0.007	63.44	63

From Tables 2 and 4, it is obvious to see that the $Std.u_i$ residuals are more dispersive than the Pearson residuals. It is really what we noted at the end of Section 4.

The results of two last examples show that how well one can provide fairly sensible answers to various questions that arise in practice for complete/incomplete contingency tables.

7 Summary

In this paper, we showed that using Relational model, one can decompose the logarithm of expected cell frequencies under a log-linear type model. This decomposition implies proper answers to several questions in the context of analyzing of complete/incomplete contingency tables. To determination of design matrix and hypothesis-induced matrix, some effective rules were given. Moreover, the properties of maximum likelihood estimators of the model parameters are obtained.

Acknowledgement

The authors would like to thank two anonymous referees for helpful comments and suggestions. The research work of S. K. Ghoreishi is supported by Qom University.

References

- Agresti, A. (2002). *Categorical Data Analysis*. Wiley, New York.
- Goodman, L.O. (1972). Some Multiplicative Models for the Analysis of Cross-Classified Data in: Proceedings of the Sixth Berkely Symposium of Mathematical Statistics and Probability.
- Goodman, L.O. (1985). The Analysis of Cross-Classified Data Having Ordered and/or Unordered Categories: Association Models, Correlation Models and Asymmetry Models for Contingency Tables with or Without Missing Entries. *The Annals of Statistics*, **13**, 10-69.

Goodman, L.O. (1991). Measures, Models, and Graphical Displays in the Analysis of Cross-Classified Data. *J. Amer. Statist. Assoc.*, **86**, 1085-1111.

Hara, H., Sei, T. and Takemura, A. (2012). Hierarchical Subspace Models for Contingency Tables. *J. of Multivariate Analysis*, **103**, 19-34.

Højsgaard, S. (2004). Statistical Inference in Context Specific Interaction Models for Contingency Tables, *Scand. J. Statist.*, **31**, 143-158.

Klimova, A., Rudas, T. and Dobra, A. (2012). Relational Models for Contingency Tables. *J. of Multivariate Analysis*, **104**, 159-173.

Lauritzen, S.I. (1996). *Graphical Models*. Oxford University Press, Oxford.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

S. K. Ghoreishi

Department of Statistics,

Qom University,

Qom, Iran.

email: atty_ghoreishi@yahoo.com

M. R. Meshkani

Department of Statistics,

Shahid Beheshti University,

Tehran, Iran.

email: m-meshkani@sbu.ac.ir