On the occurrence of boundary solutions in two-way incomplete tables

Authors: Sayan Ghosh
– Department of Mathematics, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India. (sayang@math.iitb.ac.in)

Palaniappan Vellaisamy
– Department of Mathematics, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India. (pv@math.iitb.ac.in)

Abstract:

- The analysis of incomplete contingency tables is an important problem, which is also of practical interest. In this paper, we consider boundary solutions under nonignorable nonresponse models in two-way incomplete tables with data on both variables missing. We establish a result similar to [9] on sufficient conditions for the occurrence of boundary solutions. We also provide a new result, which connects the forms of boundary solutions under various parameterizations of the missing data models. This result helps us to obtain the exact form of boundary solutions in the above tables, which improves a claim made in [2] and avoids computational burden. A counterexample is provided to show that the sufficient conditions for the occurrence of boundary solutions are not necessary, thereby disproving a conjecture of [7]. Finally, we establish new necessary conditions for the occurrence of boundary solutions under nonignorable nonresponse models in square two-way incomplete tables, and show that they are not sufficient. These conditions are simple and easy to check as they depend only on the observed cell counts. They are useful and important for model selection also. Some real life data sets are analyzed to illustrate the results.

Key-Words:

- *Incomplete tables; Boundary solutions; Log-linear models; NMAR models.*

AMS Subject Classification:

- 62H17.

1.   INTRODUCTION

Contingency tables with fully observed counts and partially classified margins (nonresponses) are called incomplete tables. The following three types of missing data mechanisms have been proposed in the literature ([8]): missing completely at random (MCAR), missing at random (MAR) and not missing at random (NMAR). The missing mechanism is said to be (a) MCAR when missingness is independent of both observed and unobserved data, (b) MAR when missingness depends only on observed data, and (c) NMAR if missingness depends on unobserved data. Nonresponses are called ignorable when the missing data mechanism is MAR or MCAR, and the parameters governing the missing data mechanism are distinct from those to be estimated. They are nonignorable when the missing data mechanism is NMAR.

Log-linear models have generally been used to study missing data mechanisms in incomplete tables (see [9] and references therein). However, under nonignorable models, a boundary solution occurs when the cell probabilities of non-respondents are estimated to be zeros for certain levels of the missing variables. That is, the maximum likelihood estimators (MLE's) of the parameters lie on the boundary of the parameter space. Note that the problem of boundary solutions is an important one as it has serious consequences for statistical inference. For example, the observed counts cannot be reproduced by a perfect fit model (a model for which the estimated expected counts are equal to the observed counts) if boundary solutions occur. This implies that the fit is inadequate and the parameter estimates are imprecise. The log likelihood function is flat and, therefore, convergence of the EM algorithm to the boundary MLE's requires a lot of iterations. Also, the eigenvalues of the covariance matrix are inappropriate (either around zero or negative), which implies some parameter estimates have large estimated standard errors and wide confidence intervals. Hence, it is useful to study various forms of boundary solutions and explore conditions for their occurrence in incomplete tables.

Consider two categorical variables with I and J levels. Then an $I \times J \times 2$ table and an $I \times J \times 2 \times 2$ table represent two-way incomplete tables with data on one of the variables and data on both the variables missing respectively. The problem of boundary solutions was first considered by [1] who proposed a sufficient condition for their occurrence in a $2 \times 2 \times 2$ incomplete table. [2] studied the problem for an $I \times J \times 2 \times 2$ incomplete table, which has non-monotone missing value patterns. For an $I \times J \times 2$ incomplete table with simple monotone missing value patterns, [10] and [3] described the problem geometrically, while [4] discussed properties of MLE's in case of boundary solutions. [9] proposed sufficient conditions for the occurrence of boundary solutions under various NMAR models in an $I \times I \times 2 \times 2$ incomplete table. Recently, [5] provided forms of boundary solutions in arbitrary three-way and n-dimensional incomplete tables with one or more variables missing, and also established sufficient conditions for their occurrence under various NMAR models. In this paper, we consider the above and

other related issues for an $I \times J \times 2 \times 2$ table. Note that a lower dimensional incomplete table is not a special case of a higher dimensional one and hence any result for the former cannot be obtained directly from that for the latter.

The purpose of this paper is to provide a comprehensive treatment of the problem of boundary solutions in two-way incomplete tables with both variables missing. To this effect, we first introduce some notations and consider various identifiable NMAR log-linear models (Models [M1]-[M5]) for an $I \times J \times 2 \times 2$ incomplete table. The problem of boundary solutions, along with their forms under the above models, is discussed in Section 3. We formally define boundary solutions for an $I \times J \times 2 \times 2$ incomplete table by extending the definition of [1], which are unavailable in the literature. A novel result (Proposition 3.1) is provided, which gives the relationship among forms of boundary solutions according to various parameterizations for the missing data models. This helps us to theoretically justify and deduce the exact boundary solutions in those models directly without having to obtain them empirically (see pp. 39-40 of [9]) using the EM algorithm. In Section 4, we illustrate this result using some data analysis examples from [2], thereby improving a claim made by them on the forms of boundary solutions in $I \times J \times 2 \times 2$ tables, which also eliminates computations.

In Section 5, we provide a result (Theorem 5.1) on sufficient conditions for the occurrence of boundary solutions in the above tables, which is similar to Theorem 1 of [9] but proved using direct arguments instead of contrapositive ones used in [9]. While [9] consider only Model [M5] in Theorem 1, we consider Models [M1]-[M5] in Theorem 5.1. A counterexample is provided to show that the sufficient conditions for the occurrence of boundary solutions are not necessary, which refutes a conjecture due to [7].

Finally, we propose new necessary conditions in Theorem 5.2 for the occurrence of boundary solutions under Models [M1]-[M5] in square two-way incomplete tables, and later show that they are not sufficient through a counterexample. Such conditions do not exist in the literature. Note that these conditions help us to identify the non-occurrence of boundary solutions, which is very useful for fitting appropriate models to the incomplete data (model selection). Also, these conditions involve only the observed cell counts and their sums in the tables, and hence can be easily verified. Section 6 provides some concluding remarks.

2. NMAR LOG-LINEAR MODELS

Suppose Y_1 and Y_2 are two categorical variables having I and J levels respectively. For $i = 1, 2$, let R_i denote the missing indicator for Y_i so that $R_i = 1$ or 2 if Y_i is observed or unobserved. Then we have an $I \times J \times 2 \times 2$ incomplete table, corresponding to Y_1, Y_2, R_1 and R_2, with cell counts $\mathbf{y} = \{y_{ijkl}\}$ where $1 \le i \le I$, $1 \le j \le J$ and $1 \le k, l \le 2$. The vector of observed counts is $\mathbf{y}_{\text{obs}} = (\{y_{ij11}\}, \{y_{i+12}\}, \{y_{+j21}\}, y_{++22})$, where $\{y_{ij11}\}$ are the fully observed

counts and $\{y_{i+12}\}, \{y_{+j21}\}, y_{++22}$ are the partially classified counts also known as the supplementary margins. All cell counts are assumed to be positive. The fully observed counts are those for which data on both Y_1 and Y_2 is available, while data on at most Y_1 or Y_2 is available for the supplementary margins. Note that '+' denotes summation over levels of the corresponding variable. For example, y_{+j21} denotes the number of observations corresponding to $Y_2 = j$ for which data on Y_2 is observed but data on Y_1 is missing. Let $\pi = \{\pi_{ijkl}\}$ be the vector of cell probabilities, $\mu = \{\mu_{ijkl}\}$ be the vector of expected counts and $N = \sum_{i,j,k,l} y_{ijkl}$ the total number of cell counts. For $I = J = 2$, we have the $2 \times 2 \times 2 \times 2$ incomplete table given by Table 1.

		$R_2 = 1$		$R_2 = 2$
		$Y_2 = 1$	$Y_2 = 2$	Y_2 missing
$R_1 = 1$	$Y_1 = 1$	y_{1111}	y_{1211}	y_{1+12}
	$Y_1 = 2$	y_{2111}	y_{2211}	y_{2+12}
$R_1 = 2$	Y_1 missing	y_{+121}	y_{+221}	y_{++22}

Table 1: $2 \times 2 \times 2 \times 2$ Incomplete Table

We consider Poisson sampling for convenience, that is, $Y_{ijkl} \sim P(\mu_{ijkl})$. The likelihood function of μ is

$$(2.1) \qquad L(\mu; \mathbf{y}_{\text{obs}}) = \frac{e^{-\sum_{i,j,k,l} \mu_{ijkl}} \prod_{i,j} \mu_{ij11}^{y_{ij11}} \prod_i \mu_{i+12}^{y_{i+12}} \prod_j \mu_{+j21}^{y_{+j21}} \mu_{++22}^{y_{++22}}}{\prod_{i,j,k,l} y_{ijkl}!}$$

so that the log-likelihood function of μ is

$$l(\mu; \mathbf{y}_{\text{obs}}) = \sum_{i,j} y_{ij11} \log \mu_{ij11} + \sum_i y_{i+12} \log \mu_{i+12} + \sum_j y_{+j21} \log \mu_{+j21}$$

$$(2.2) \qquad + y_{++22} \log \mu_{++22} - \mu_{++++} + \Delta,$$

where Δ is independent of μ_{ijkl}'s. For an $I \times J \times 2 \times 2$ incomplete table, [2] proposed the following log-linear model (with no three-way or four-way interactions):

$$\log \mu_{ijkl} = \lambda + \lambda_{Y_1}(i) + \lambda_{Y_2}(j) + \lambda_{R_1}(k) + \lambda_{R_2}(l) + \lambda_{Y_1 Y_2}(i, j)$$

$$(2.3) \qquad + \lambda_{Y_1 R_1}(i, k) + \lambda_{Y_2 R_1}(j, k) + \lambda_{Y_1 R_2}(i, l) + \lambda_{Y_2 R_2}(j, l) + \lambda_{R_1 R_2}(k, l),$$

where the sum over any argument of a log-linear parameter is zero, for example, $\sum_i \lambda_{Y_1 Y_2}(i, j) = \sum_j \lambda_{Y_1 Y_2}(i, j) = 0$. To study the various missing mechanisms of Y_1 and Y_2, [2] introduced the following notations:

$$a_{ij} = \frac{P(R_1 = 2, R_2 = 1 | Y_1 = i, Y_2 = j)}{P(R_1 = 1, R_2 = 1 | Y_1 = i, Y_2 = j)} = \frac{\pi_{ij21}}{\pi_{ij11}} = \frac{\mu_{ij21}}{\mu_{ij11}},$$

$$b_{ij} = \frac{P(R_1 = 1, R_2 = 2 | Y_1 = i, Y_2 = j)}{P(R_1 = 1, R_2 = 1 | Y_1 = i, Y_2 = j)} = \frac{\pi_{ij12}}{\pi_{ij11}} = \frac{\mu_{ij12}}{\mu_{ij11}}, \quad m_{ij11} = N\pi_{ij11},$$

$$g = \frac{P(R_1 = 1, R_2 = 1 | Y_1 = i, Y_2 = j)P(R_1 = 2, R_2 = 2 | Y_1 = i, Y_2 = j)}{P(R_1 = 1, R_2 = 2 | Y_1 = i, Y_2 = j)P(R_1 = 2, R_2 = 1 | Y_1 = i, Y_2 = j)}.$$

Remark 2.1.   Under (2.3), it can be shown that $a_{ij} = \exp[-2\{\lambda_{R_1}(1) + \lambda_{Y_1 R_1}(i,1) + \lambda_{Y_2 R_1}(j,1) + \lambda_{R_1 R_2}(1,1)\}]$ and $b_{ij} = \exp[-2\{\lambda_{R_2}(1) + \lambda_{Y_1 R_2}(i,1) + \lambda_{Y_2 R_2}(j,1) + \lambda_{R_1 R_2}(1,1)\}]$. Also, we have $g = \frac{\pi_{ij11}\pi_{ij22}}{\pi_{ij12}\pi_{ij21}} = \frac{\mu_{ij11}\mu_{ij22}}{\mu_{ij12}\mu_{ij21}}$. Hence

$$\log g = \log \mu_{ij11} + \log \mu_{ij22} - \log \mu_{ij12} - \log \mu_{ij21}$$
$$\Rightarrow \log g = \lambda_{R_1 R_2}(1,1) + \lambda_{R_1 R_2}(2,2) - \lambda_{R_1 R_2}(1,2) - \lambda_{R_1 R_2}(2,1) \quad \text{(from (2.3))}$$
$$\Rightarrow \log g = 4\lambda_{R_1 R_2}(1,1) \quad (\because \lambda_{R_1 R_2}(1,2) = -\lambda_{R_1 R_2}(2,2) = \lambda_{R_1 R_2}(2,1)$$
$$= -\lambda_{R_1 R_2}(1,1))$$
$$\Rightarrow g = \exp[4\lambda_{R_1 R_2}(1,1)],$$

which is independent of i and j.

Note that $m_{ij11} = \mu_{ij11}$ and g denotes the odds ratio between the missing indicators of Y_1 and Y_2. Also, $\mu_{ij21} = m_{ij11}a_{ij}$, $\mu_{ij12} = m_{ij11}b_{ij}$ and $\mu_{ij22} = m_{ij11}a_{ij}b_{ij}g$. Note that a_{ij} is the conditional odds of Y_1 being missing given Y_2 is observed, while b_{ij} is the conditional odds of Y_2 being missing given Y_1 is observed. Here, a_{ij} and b_{ij} describe the missing mechanisms of Y_1 and Y_2, respectively. Denote a_{ij} (b_{ij}) by $\alpha_{i.}$ ($\beta_{i.}$) or $\alpha_{.j}$ ($\beta_{.j}$) or $\alpha_{..}$ ($\beta_{..}$) if it depends only on i or j or none, respectively. Then we have the following definition.

Definition 2.1.   The missing mechanism of Y_1 under (2.3) is NMAR if $a_{ij} = \alpha_{i.}$, MAR if $a_{ij} = \alpha_{.j}$ and MCAR if $a_{ij} = \alpha_{..}$. Similarly, the missing mechanism of Y_2 is NMAR if $b_{ij} = \beta_{.j}$, MAR if $b_{ij} = \beta_{i.}$ and MCAR if $b_{ij} = \beta_{..}$.

Using Definition 2.1 and the above notations, there are nine possible identifiable models (see pp. 647-648 of [2]) based on different missing mechanisms for Y_1 and Y_2. The equivalent log-linear models can be obtained as submodels of (2.3). As an example, consider the model $(\alpha_{i.}, \beta_{i.})$, for which the missing mechanism is NMAR for Y_1 and MAR for Y_2. Using the expressions of a_{ij} and b_{ij} in Remark 2.1, the corresponding log-linear model is obtained from (2.3) by substituting $\lambda_{Y_2 R_1}(j,k) = \lambda_{Y_2 R_2}(j,l) = 0$. The following are the five models when the missing mechanism is NMAR for Y_1 or Y_2.

1. Model M1 (NMAR for Y_1, MCAR for Y_2):

$$\log \mu_{ijkl} = \lambda + \lambda_{Y_1}(i) + \lambda_{Y_2}(j) + \lambda_{R_1}(k) + \lambda_{R_2}(l) + \lambda_{Y_1 Y_2}(i,j) + \lambda_{Y_1 R_1}(i,k) + \lambda_{R_1 R_2}(k,l)$$

2. Model M2 (NMAR for Y_2, MCAR for Y_1):

$$\log \mu_{ijkl} = \lambda + \lambda_{Y_1}(i) + \lambda_{Y_2}(j) + \lambda_{R_1}(k) + \lambda_{R_2}(l) + \lambda_{Y_1 Y_2}(i,j) + \lambda_{Y_2 R_2}(j,l) + \lambda_{R_1 R_2}(k,l)$$

3. Model M3 (NMAR for Y_1, MAR for Y_2):

$$\begin{aligned}\log \mu_{ijkl} = {} & \lambda + \lambda_{Y_1}(i) + \lambda_{Y_2}(j) + \lambda_{R_1}(k) + \lambda_{R_2}(l) + \lambda_{Y_1 Y_2}(i,j) + \lambda_{Y_1 R_1}(i,k) \\ & + \lambda_{Y_1 R_2}(i,l) + \lambda_{R_1 R_2}(k,l)\end{aligned}$$

4. Model M4 (NMAR for Y_2, MAR for Y_1):

$$\begin{aligned}
\log \mu_{ijkl} = & \lambda + \lambda_{Y_1}(i) + \lambda_{Y_2}(j) + \lambda_{R_1}(k) + \lambda_{R_2}(l) + \lambda_{Y_1 Y_2}(i,j) + \lambda_{Y_2 R_1}(j,k) \\
& + \lambda_{Y_2 R_2}(j,l) + \lambda_{R_1 R_2}(k,l)
\end{aligned}$$

5. Model M5 (NMAR for both Y_1 and Y_2):

$$\begin{aligned}
\log \mu_{ijkl} = & \lambda + \lambda_{Y_1}(i) + \lambda_{Y_2}(j) + \lambda_{R_1}(k) + \lambda_{R_2}(l) + \lambda_{Y_1 Y_2}(i,j) + \lambda_{Y_1 R_1}(i,k) \\
& + \lambda_{Y_2 R_2}(j,l) + \lambda_{R_1 R_2}(k,l)
\end{aligned}$$

Note that for Models [M1]-[M5], there is an association term between a variable and its missing indicator if the missing mechanism is NMAR for that variable (for example, the term $\lambda_{Y_1 R_1}(i,k)$ in Model [M1]), between a variable and the other missing indicator if the missing mechanism is MAR for that variable (for example, the term $\lambda_{Y_2 R_1}(j,k)$ in Model [M4]) and none if the missing mechanism is MCAR for a variable (for example, $\lambda_{Y_1 R_1}(i,k)$ and $\lambda_{Y_2 R_1}(j,k)$ are absent in Model [M2]).

3. BOUNDARY SOLUTIONS IN NMAR MODELS

In this section, we consider boundary solutions under non-ignorable nonresponse (NMAR) models for an $I \times J \times 2 \times 2$ incomplete table. We first define boundary solutions under the above models and then present a result relating the forms of boundary solutions in terms of various parameterizations of the models.

For an incomplete table, boundary solutions in NMAR models occur when the MLE's of nonresponse cell probabilities are all zeros for certain levels of the missing variables. For an $I \times J \times 2$ incomplete table, where data on only Y_2 is missing, [1] defined boundary solutions in the NMAR model for Y_2 as $\hat{\pi}_{ij2} = 0$ for at least one pair (i,j). For the same model, [4] showed that boundary solutions are given by $\hat{\pi}_{+j2} = 0$ for at least one and at most $(J-1)$ values of Y_2. [1] defined a nonresponse boundary solution under NMAR models in general to be a stationary point that lies on a boundary of the space of parameters modeling the nonignorable nonresponse. Using this, we may extend their definition to an $I \times J \times 2 \times 2$ table as follows.

Definition 3.1.   Consider an $I \times J \times 2 \times 2$ incomplete table, and let $1 \leq i \leq I$, $1 \leq j \leq J$ and $k, l = 1, 2$. Then we have the following.
1. A nonresponse boundary solution under the NMAR models for Y_1 only, that is, Models [M1] and [M3] is an MLE given by $\hat{\pi}_{ij2l} = 0$ for at least one combination (i,j,l).
2. A nonresponse boundary solution under the NMAR models for Y_2 only, that is, Models [M2] and [M4] is an MLE given by $\hat{\pi}_{ijk2} = 0$ for at least one combination (i,j,k).
3. A nonresponse boundary solution under the NMAR model for both Y_1 and

Y_2, that is, Model [M5] is an MLE given by $\hat{\pi}_{ij2l} = 0$ for at least one combination (i, j, l) or $\hat{\pi}_{ijk2} = 0$ for at least one combination (i, j, k).

Note that in the literature, boundary solutions have usually been defined in terms of cell probabilities because the cell probabilities are in some sense natural to the model for the incomplete table, whereas the loglinear parameters are not. The next proposition explores the relationships among boundary solutions under Models [M1]-[M5] in terms of MLE's of nonresponse cell probabilities, some specific log-linear parameters and $\alpha_{i.}$ or $\beta_{.j}$ for two-way incomplete tables with both variables missing.

Proposition 3.1. *For an $I \times J \times 2 \times 2$ incomplete table, we have the following.*
1. For Models [M1] and [M3], if boundary solutions occur, then they are given by $\hat{\lambda}_{Y_1 R_1}(i, 2) = -\infty \Leftrightarrow \hat{\pi}_{i+2+} = 0 \Leftrightarrow \hat{\alpha}_{i.} = 0$ for at least one and at most $(I - 1)$ values of Y_1.
2. For Models [M2] and [M4], if boundary solutions occur, then they are given by $\hat{\lambda}_{Y_2 R_2}(j, 2) = -\infty \Leftrightarrow \hat{\pi}_{+j+2} = 0 \Leftrightarrow \hat{\beta}_{.j} = 0$ for at least one and at most $(J - 1)$ values of Y_2.
3. For Model [M5], if boundary solutions occur, then they are given by $\hat{\lambda}_{Y_1 R_1}(i, 2) = -\infty$ or $\hat{\lambda}_{Y_2 R_2}(j, 2) = -\infty \Leftrightarrow \hat{\pi}_{i+2+} = 0$ or $\hat{\pi}_{+j+2} = 0 \Leftrightarrow \hat{\alpha}_{i.} = 0$ for at least one and at most $(I - 1)$ values of Y_1 or $\hat{\beta}_{.j} = 0$ for at least one and at most $(J - 1)$ values of Y_2.

Proof: See Appendix A1. $\square$

From the proof of Proposition 3.1 in Appendix A1, note that the one-to-one relation between the cell probabilities and the log-linear parameters cannot be used to derive the connection between the different forms of boundary solutions. This is because it is not obvious which specific log-linear parameters have infinite MLE's just by noting the zero MLE's of the nonresponse cell probabilities when boundary solutions occur.

4.   SOME EXAMPLES OF BOUNDARY SOLUTIONS IN NMAR MODELS

In this section, we reanalyze some examples in [2], illustrating the result in Section 3. We use Proposition 3.1 to investigate a claim made by [2] regarding forms and occurrence of boundary solutions in an $I \times J \times 2 \times 2$ incomplete table. This improvement is useful as it avoids computation and provides the exact boundary solutions under a NMAR model by simply noting the level (s) of the variable (s) for which the MLE's of the parameters are negative or infinite.

First, we present the correct expression of the likelihood ratio statistic for missing data models in such a table. Consider testing the goodness of fit of a null model (here one of the Models [M1]-[M5]) against the alternative model (perfect fit model). Let $\{\hat{\mu}_{ijkl}\}$ and $\{\tilde{\mu}_{ijkl}\}$ denote the MLE's of the expected counts under a null model and a perfect fit model respectively. Also, let L_0 and L_1 denote the log-likelihoods for the null and the alternative models, respectively. Then the likelihood ratio statistic is given by

$$
\begin{aligned}
G^2 &= -2(L_0 - L_1) \\
&= -2\left[\sum_{i,j} y_{ij11} \ln\left(\frac{\hat{\mu}_{ij11}}{\tilde{\mu}_{ij11}}\right) + \sum_i y_{i+12} \ln\left(\frac{\hat{\mu}_{i+12}}{\tilde{\mu}_{i+12}}\right) + \sum_j y_{+j21} \ln\left(\frac{\hat{\mu}_{+j21}}{\tilde{\mu}_{+j21}}\right) \right. \\
&\quad \left. + y_{++22} \ln\left(\frac{\hat{\mu}_{++22}}{\tilde{\mu}_{++22}}\right) - \hat{\mu}_{++++} + \tilde{\mu}_{++++} \right] \\
&= -2\left[\sum_{i,j} y_{ij11} \ln\left(\frac{\hat{m}_{ij11}}{y_{ij11}}\right) + \sum_i y_{i+12} \ln\left(\frac{\sum_j \hat{m}_{ij11}\hat{b}_{ij}}{y_{i+12}}\right) \right. \\
&\quad + \sum_j y_{+j21} \ln\left(\frac{\sum_i \hat{m}_{ij11}\hat{a}_{ij}}{y_{+j21}}\right) + y_{++22} \ln\left(\frac{\sum_{i,j} \hat{m}_{ij11}\hat{a}_{ij}\hat{b}_{ij}\hat{g}}{y_{++22}}\right) \\
(4.1) &\quad \left. - \sum_{i,j} \hat{m}_{ij11}(1 + \hat{a}_{ij} + \hat{b}_{ij} + \hat{a}_{ij}\hat{b}_{ij}\hat{g}) + N \right].
\end{aligned}
$$

Note that the last two terms of (4.1) are missing in the expression of G^2 in [2] (see p. 646). Observe that in general, $\sum_{i,j} \hat{m}_{ij11}(1 + \hat{a}_{ij} + \hat{b}_{ij} + \hat{a}_{ij}\hat{b}_{ij}\hat{g}) \neq N$, unless the hypothetical (null) model is a perfect fit model for example, in which case $G^2 = 0$.

Using Definition 2.1 and the notations in Section 2, Models [M1]-[M5] can be represented as follows − Model [M1]: $(\alpha_{i.}, \beta_{..})$, Model [M2]: $(\alpha_{..}, \beta_{.j})$, Model [M3]: $(\alpha_{i.}, \beta_{i.})$, Model [M4]: $(\alpha_{.j}, \beta_{.j})$ and Model [M5]: $(\alpha_{i.}, \beta_{.j})$. Accordingly, the expression of G^2 in (4.1) for each of the above models may be obtained by making suitable substitutions and using the MLE's in [2] (see pp. 647-648). For example, the MLE's under the model $(\alpha_{i.}, \beta_{..})$ are

$$
\hat{m}_{ij11} = \frac{y_{ij11}y_{i+1+}y_{++11}}{y_{i+11}y_{++1+}}, \quad \sum_i \hat{m}_{ij11}\hat{\alpha}_{i.} = y_{+j21}, \quad \hat{\beta}_{..} = \frac{y_{++12}}{y_{++11}}, \quad \hat{g} = \frac{y_{++11}y_{++22}}{y_{++12}y_{++21}}.
$$

Hence, from (4.1), the likelihood ratio statistic is

$$
G^2 = -2\left[\sum_{i,j} y_{ij11} \ln\left(\frac{y_{i+1+}y_{++11}}{y_{i+11}y_{++1+}}\right) + \sum_i y_{i+12} \ln\left(\frac{y_{i+1+}y_{++12}}{y_{i+12}y_{++1+}}\right) \right].
$$

[2] mentioned that if any solution $\hat{\alpha}_{i.}$ or $\hat{\beta}_{.j}$ to the systems of equations $\sum_i N\hat{\pi}_{ij11}\hat{\alpha}_{i.} = y_{+j21}$ and $\sum_j N\hat{\pi}_{ij11}\hat{\beta}_{.j} = y_{i+12}$ respectively is negative, then

boundary solutions occur, that is, the MLE lies on the boundary of the param-
eter space. Closed-form boundary MLE's under Models [M1]-[M5] may then be
obtained (see p. 649 of [2]) by setting certain parameter estimates ($\hat{\alpha}_{i.}$ or $\hat{\beta}_{.j}$) to 0
in the likelihood equations obtained from (2.2) for the models. They claimed that
counterintuitively, the parameter estimate set to 0 need not be the estimate with
a negative value as the solution to the above systems of equations. In particular,
for a $2 \times 2 \times 2 \times 2$ incomplete table, they suggested examining both boundaries
$\hat{\alpha}_{1.} = 0$ and $\hat{\alpha}_{2.} = 0$; similarly $\hat{\beta}_{.1} = 0$ and $\hat{\beta}_{.2} = 0$ to determine the minimum
value of G^2, which corresponds to the MLE. We improve this claim and thereby
obviate computations by showing that the MLE indeed always occurs on the spe-
cific boundary (level (s) of the variable (s)) for which $\hat{\alpha}_{i.}$ or $\hat{\beta}_{.j}$ is negative. In the
next three examples, we use Proposition 3.1 to illustrate this point for Models
[M1]-[M5].

Example 4.1.   Consider the data in Table 2 discussed in [2], which cross-
classifies mother's self-reported smoking status (Y_1) ($Y_1 = 1(2)$ for smoker (non-
smoker)) with newborn's weight (Y_2) ($Y_2 = 1(2)$ if weight < 2500 grams (≥ 2500
grams)). The supplementary margins contain data on only smoking status, data
on only newborn's weight and missing data on both variables.

		$R_2 = 1$		$R_2 = 2$
		$Y_2 = 1$	$Y_2 = 2$	Y_2 missing
$R_1 = 1$	$Y_1 = 1$	4512	21009	1049
	$Y_1 = 2$	3394	24132	1135
$R_1 = 2$	Y_1 missing	142	464	1224

Table 2:   Birth weight and smoking: observed counts

[2] mentioned that $\hat{\alpha}_{2.} < 0$ is obtained on fitting models [M1], [M3] and [M5]
to the data in Table 2. Also, the value of G^2 corresponding to $\hat{\alpha}_{2.} = 0$ is larger
than that corresponding to $\hat{\alpha}_{1.} = 0$ for all the above models, which is incorrect
as shown below. When we fit the same models to the data in Table 2 using the
'MASS' package in R software, we obtain $\hat{\alpha}_{1.} = 0.0493$ and $\hat{\alpha}_{2.} = -0.0237$ under
Models [M1], [M3] and [M5], that is, boundary solutions occur in each of the
models.

Also, $G^2 = 55.2198$ (12.4682) under Model [M1], $G^2 = 55.2168$ (12.4638)
under Model [M3] and $G^2 = 55.214$ (12.464) under Model [M5] when $\hat{\alpha}_{1.} =
0$ ($\hat{\alpha}_{2.} = 0$). The G^2 values for $\hat{\alpha}_{2.} = 0$ upon rounding off in each of the models
match those given in Table V of [2]. Hence, G^2 is minimum for $\hat{\alpha}_{2.} = 0$ in each
case, which implies that boundary solutions are given by $\hat{\alpha}_{2.} = 0$ or equivalently
$\hat{\pi}_{2+2+} = 0$. This result is consistent with points 1 and 3 of Proposition 3.1.
Further, it is the exact form of boundary solutions that we obtain on fitting
Models [M1], [M3] and [M5] to the data in Table 2 using the EM algorithm (see
the 'ecm.cat' function of 'cat' package in R software).

Example 4.2. Consider the example given in the last paragraph of p. 646 in [2]. The model [M1] was fitted to the following data: $y_{1111} = 100$, $y_{1211} = 40$, $y_{2111} = 50$, $y_{2211} = 1000$, $y_{1+12} = 0$, $y_{2+12} = 0$, $y_{+121} = 100$, $y_{+221} = 10$ and $y_{++22} = 0$. They mentioned that though $\hat{\alpha}_{1.} < 0$, G^2 is minimum for $\hat{\alpha}_{2.} = 0$ implying that the MLE is on the boundary $\hat{\alpha}_{2.} = 0$. However, we obtain $\hat{\alpha}_{1.} = 1.0153$ (> 0) and $\hat{\alpha}_{2.} = -0.0306$ on fitting Model [M1] to the above data. Also, note that $\hat{g} = \frac{y_{++11}y_{++22}}{y_{++12}y_{++21}}$ (see p. 649 of [2]) is undefined since $y_{++12} = 0$. Hence, we introduce the following changes: $y_{1+12} = 1$, $y_{2+12} = 1$ and $y_{++22} = 2$ as shown in Table 3.

		$R_2 = 1$		$R_2 = 2$
		$Y_2 = 1$	$Y_2 = 2$	Y_2 missing
$R_1 = 1$	$Y_1 = 1$	100	40	1
	$Y_1 = 2$	50	1000	1
$R_1 = 2$	Y_1 missing	100	10	2

Table 3: Modified $2 \times 2 \times 2 \times 2$ table

On fitting models [M1], [M3] and [M5] to the data in Table 3, we obtain $\hat{\alpha}_{1.} = 1.0098$ under [M1], and $\hat{\alpha}_{1.} = 1.0153$ under [M3] and [M5], along with $\hat{\alpha}_{2.} = -0.0306$ under all the above models, which implies boundary solutions occur in each case. Also, $G^2 = 426.1604$ (17.4704) under Model [M1], $G^2 = 424.3288$ (15.669) under Model [M3] and $G^2 = 424.3188$ (15.664) under Model [M5] when $\hat{\alpha}_{1.} = 0$ ($\hat{\alpha}_{2.} = 0$). Hence, G^2 is minimum for $\hat{\alpha}_{2.} = 0$ in each model, which implies that boundary solutions are given by $\hat{\pi}_{2+2+} = 0$. This result is consistent with points 1 and 3 of Proposition 3.1. Further, it is the exact form of boundary solutions that we obtain on fitting Models [M1], [M3] and [M5] to the data in Table 3 using the EM algorithm.

Example 4.3. Consider the data in Table 2 discussed in Example 4.1. We introduce the following changes corresponding to supplementary margins in Table 2: $464 \rightarrow 700$ and $1135 \rightarrow 750$. The modified table is shown in Table 4.

		$R_2 = 1$		$R_2 = 2$
		$Y_2 = 1$	$Y_2 = 2$	Y_2 missing
$R_1 = 1$	$Y_1 = 1$	4512	21009	1049
	$Y_1 = 2$	3394	24132	750
$R_1 = 2$	Y_1 missing	142	700	1224

Table 4: Birth weight and smoking: observed counts (modified)

When we fit the models [M2], [M4] and [M5] to the data in Table 4, we

obtain $\hat{\beta}_{.1} = 0.2538$ under [M2], and $\hat{\beta}_{.1} = 0.2543$ under [M4] and [M5] along with $\hat{\beta}_{.2} = -0.0047$ under all the above models, that is, boundary solutions occur in each of the models. Also, $G^2 = 98.5962$ (3.3548) under Model [M2], $G^2 = 96.1622$ (0.922) under Model [M4] and $G^2 = 96.162$ (0.9276) under Model [M5] when $\hat{\beta}_{.1} = 0$ ($\hat{\beta}_{.2} = 0$). The G^2 values in brackets above match those obtained using the EM algorithm. Hence, G^2 is minimum for $\hat{\beta}_{.2} = 0$ in each case, which implies that boundary solutions are given by $\hat{\beta}_{.2} = 0$ or equivalently $\hat{\pi}_{+2+2} = 0$. This result is consistent with points 2 and 3 of Proposition 3.1. Further, it is the exact form of boundary solutions that we obtain on fitting Models [M2], [M4] and [M5] to the data in Table 4 using the EM algorithm.

5.   CONDITIONS FOR THE OCCURRENCE OF BOUNDARY SOLUTIONS

In this section, we discuss sufficient conditions and also propose necessary conditions for the occurrence of boundary solutions in two-way incomplete tables with both variables missing. We show that the sufficient conditions are not necessary, which disproves a conjecture made by [7]. Further, we prove that the proposed necessary conditions are not sufficient. Both sets of conditions are simple to verify since they involve only the observed cell counts in the tables. The sufficient conditions and the necessary conditions are of practical utility in identifying the occurrence and non-occurrence, respectively of boundary solutions in such tables.

5.1. Sufficient conditions for the occurrence of boundary solutions

Following Park *et al.* (2014), define the four odds based on the observed (joint/marginal) cell counts for any pair (j, j') of Y_2:

$$\nu_i(j, j') = \frac{\hat{\pi}_{ij11}}{\hat{\pi}_{ij'11}}, \ \nu_n(j, j') = \min_i\{\nu_i(j, j')\}, \ \nu_m(j, j') = \max_i\{\nu_i(j, j')\},$$

$$(5.1) \quad \nu(j, j') = \frac{y_{+j21}}{y_{+j'21}}.$$

Similarly, for a given pair (i, i') of Y_1, define the four odds using the observed cell counts:

$$\omega_j(i, i') = \frac{\hat{\pi}_{ij11}}{\hat{\pi}_{i'j11}}, \ \omega_n(i, i') = \min_j\{\omega_j(i, i')\}, \ \omega_m(i, i') = \max_j\{\omega_j(i, i')\},$$

$$(5.2) \quad \omega(i, i') = \frac{y_{i+12}}{y_{i'+12}}.$$

Note that $\nu_i(j, j')$ and $\omega_j(i, i')$ are called the response odds, while $\nu(j, j')$ and $\omega(i, i')$ are called the nonresponse odds. Using the MLE's of $\{\pi_{ij11}\}$ under Models

[M1]-[M5] (see pp. 647-648 of [2]), we deduce that $\nu_i(j, j') = \frac{y_{ij11}}{y_{ij'11}}$ and $\omega_j(i, i') = \frac{y_{ij11}}{y_{i'j11}}$, which involve only the fully observed counts.

Theorem 1 of [9] deals with sufficient conditions for the occurrence of boundary solutions only under Model [M5]. However, in the next result, we provide such conditions for the occurrence of boundary solutions under Models [M1]-[M5]. Also, we provide a proof which is similar to that of Theorem 1 of [9], but we give direct arguments, which are different from the contrapositive ones used by [9].

Theorem 5.1. *Consider the following conditions for an $I \times I \times 2 \times 2$ contingency table.*

1. $\nu(j, j') \notin (\nu_n(j, j'), \nu_m(j, j'))$ *for at least one pair (j, j') of Y_2,*

2. $\omega(i, i') \notin (\omega_n(i, i'), \omega_m(i, i'))$ *for at least one pair (i, i') of Y_1.*

Then we have the following:

(a) *Boundary solutions in NMAR models for only Y_1 (Models [M1] and [M3]) occur if Condition 1 holds.*

(b) *Boundary solutions in NMAR models for only Y_2 (Models [M2] and [M4]) occur if Condition 2 holds.*

(c) *Boundary solutions in the NMAR model for both Y_1 and Y_2 (Model [M5]) occur if Condition 1 or Condition 2 holds.*

Proof: See Appendix A2. □

5.2. The sufficient conditions are not necessary

The next example shows that the sufficient conditions for the occurrence of boundary solutions mentioned in Theorem 5.1 are not necessary. This result has not been discussed in the literature earlier. In fact, [7] proved that the above conditions are both necessary and sufficient for a $2 \times 2 \times 2 \times 2$ incomplete table. They conjectured that a similar result would hold for general two-way incomplete tables as well.

Example 5.1. Consider Table 5 discussed in [9], which cross-classifies data on bone mineral density (Y_1) and family income (Y_2) in a $3 \times 3 \times 2 \times 2$ incomplete table. Both variables Y_1 and Y_2 have three levels. The total count is 2998 out of which data on Y_1 and Y_2 are available for 1844 persons, data on

		$R_2 = 1$			$R_2 = 2$
		$Y_2 = 1$	$Y_2 = 2$	$Y_2 = 3$	Missing
	$Y_1 = 1$	621	290	284	135
$R_1 = 1$	$Y_1 = 2$	260	131	117	69
	$Y_1 = 3$	93	30	18	27
$R_1 = 2$	Missing	456	156	266	45

Table 5:   Bone mineral density (Y_1) and family income (Y_2)

Y_1 only for 231 persons, data on Y_2 only for 878 persons, and data on neither of them for 45 persons.

Now, we introduce the following changes corresponding to supplementary margins in Table 5: $266 \to 125$, $69 \to 60$ and $27 \to 20$. The modified table is shown in Table 6.

		$R_2 = 1$			$R_2 = 2$
		$Y_2 = 1$	$Y_2 = 2$	$Y_2 = 3$	Missing
	$Y_1 = 1$	621	290	284	135
$R_1 = 1$	$Y_1 = 2$	260	131	117	60
	$Y_1 = 3$	93	30	18	20
$R_1 = 2$	Missing	456	156	125	45

Table 6:   Modified Table 5

From Table 6, $\nu(1,2) = 456/156 = 2.92$, $\nu(1,3) = 456/125 = 3.65$, $\nu(2,3) = 156/125 = 1.25$, $\omega(1,2) = 135/60 = 2.25$, $\omega(1,3) = 135/20 = 6.75$ and $\omega(2,3) = 60/20 = 3.00$. Let $I_\nu(j,j') = (\nu_n(j,j'), \nu_m(j,j'))$ and $I_\omega(i,i') = (\omega_n(i,i'), \omega_m(i,i'))$. Then from Table 6, it can be shown that $\nu(1,2) \in I_\nu(1,2) = (260/131, 93/30)$, $\nu(1,3) \in I_\nu(1,3) = (621/284, 93/18)$, $\nu(2,3) \in I_\nu(2,3) = (290/284, 30/18)$, $\omega(1,2) \in I_\omega(1,2) = (290/131, 284/117)$, $\omega(1,3) \in I_\omega(1,3) = (621/93, 284/18)$ and $\omega(2,3) \in I_\omega(2,3) = (260/93, 117/18)$ so that the sufficient conditions for the occurrence of boundary solutions in Theorem 5.1 are not satisfied. The MLE's of the parameters obtained on fitting Models [M1]-[M5] in various subtables of Table 6 are shown in Table 7.

From Table 7, note that in each subtable, at least one of $\hat{\alpha}_{i.}$ and $\hat{\beta}_{.j}$ is negative, which imply that boundary solutions occur. The forms of boundary solutions under the Models [M1]-[M5] are also the same as described in Section 3. This shows that for an $I \times J \times 2 \times 2$ incomplete table, where $I, J \geq 3$, the sufficient conditions for the occurrence of boundary solutions under Models [M1]-[M5] in Theorem 5.1 are not necessary.

Subtable	NMAR model	MLE's	Boundary solutions
Y_1	[M1]	$\hat{\alpha}_{1.} = 0.6556, \hat{\alpha}_{2.} = -1.0537, \hat{\alpha}_{3.} = 3.4109$	$\hat{\pi}_{2+2+} = 0$
Y_2	[M2]	$\hat{\beta}_{.1} = 0.1355, \hat{\beta}_{.2} = 0.3420, \hat{\beta}_{.3} = -0.1846$	$\hat{\pi}_{+3+2} = 0$
$Y_1 Y_2$	[M1]	$\hat{\alpha}_{1.} = 0.6556, \hat{\alpha}_{2.} = -1.0537, \hat{\alpha}_{3.} = 3.4109$	$\hat{\pi}_{2+2+} = 0$
	[M3]	$\hat{\alpha}_{1.} = 0.6534, \hat{\alpha}_{2.} = -1.0551, \hat{\alpha}_{3.} = 3.4874$	$\hat{\pi}_{2+2+} = 0$
$Y_1 Y_2$	[M2]	$\hat{\beta}_{.1} = 0.1355, \hat{\beta}_{.2} = 0.3420, \hat{\beta}_{.3} = -0.1846$	$\hat{\pi}_{+3+2} = 0$
	[M4]	$\hat{\beta}_{.1} = 0.1421, \hat{\beta}_{.2} = 0.3289, \hat{\beta}_{.3} = -0.1712$	$\hat{\pi}_{+3+2} = 0$
$Y_1 Y_2$	[M5]	$\hat{\alpha}_{1.} = 0.6534, \hat{\alpha}_{2.} = -1.0551, \hat{\alpha}_{3.} = 3.4874,$	$\hat{\pi}_{2+2+} = 0,$
		$\hat{\beta}_{.1} = 0.1421, \hat{\beta}_{.2} = 0.3289, \hat{\beta}_{.3} = -0.1712$	$\hat{\pi}_{+3+2} = 0$

Table 7: MLE's of parameters in subtables of Table 6

5.3. Necessary conditions for the occurrence of boundary solutions
---

We next state below a result due to [6], which will be used later to obtain a result on the occurrence of boundary solutions.

Lemma 5.1. *Suppose $A = (a_{ij})$ is a matrix with $a_{ij} \geq 0$ for $i \neq j = 1, 2, \ldots, n$ and $a_{ii} > 0$. Also, let $\mathbf{b} = (b_j)$, where $b_j > 0$ for $1 \leq j \leq n$. If*

$$(5.3) \qquad b_i > \sum_{j \neq i=1}^{n} a_{ij} \frac{b_j}{a_{jj}}, \quad \forall\, 1 \leq i \leq n,$$

then A is invertible and $A^{-1}\mathbf{b} > \mathbf{0}$.

Using Lemma 5.1, the next result provides necessary conditions for the occurrence of boundary solutions under Models [M1]-[M5] in square two-way incomplete tables.

Theorem 5.2. *For an $I \times I \times 2 \times 2$ incomplete table, consider the following conditions:*

1. $y_{+j21} \leq \sum_{i \neq j=1}^{I} \hat{\mu}_{ji11} \frac{y_{+i21}}{\hat{\mu}_{ii11}}$ *for at least one $j = 1, 2, \ldots, I$,*

2. $y_{i+12} \leq \sum_{j \neq i=1}^{I} \hat{\mu}_{ij11} \frac{y_{j+12}}{\hat{\mu}_{jj11}}$ *for at least one $i = 1, 2, \ldots, I$,*

where $\hat{\mu}_{ij11}$ is the MLE of μ_{ij11}. Also, let $\{\hat{\mu}_{ij11}\} > 0$, $\{y_{i+12}\} > 0$ and $\{y_{+j21}\} > 0$. Then we have the following:

(a) *If boundary solutions under Models [M1] and [M3] occur, then only Condition 1 holds.*

(b)   *If boundary solutions under Models [M2] and [M4] occur, then only Condition 2 holds.*

(c)   *If boundary solutions under the Model [M5] occur, then Condition 1 or Condition 2 holds.*

Proof: See Appendix A3.   □

Henceforth, we denote $A = (a_{ij}) = (\hat{\mu}_{ij11})$, $\mathbf{b} = (b_j) = (y_{+j21})$ and $\mathbf{b}^* = (b_i^*) = (y_{i+12})$ for $1 \leq i \leq I$, $1 \leq j \leq I$. The example below is an application of Theorem 5.2.

Example 5.2.   From Table 6 in Example 5.1, we have the following:

$$A = \begin{pmatrix} 621 & 290 & 284 \\ 260 & 131 & 117 \\ 93 & 30 & 18 \end{pmatrix}, \quad \mathbf{b} = (456, 156, 125), \quad \mathbf{b}^* = (135, 60, 20).$$

The MLE's $\hat{\alpha} = (\hat{\alpha}_{i.})$ and $\hat{\beta} = (\hat{\beta}_{.j})$ under Model [M5] satisfy respectively the systems $A^T \alpha = \mathbf{b}$ from (6.11) and $A\beta = \mathbf{b}^*$ from (6.12) for $i, j = 1, 2, 3$. From Table 7, we observe that if Model [M5] is fitted to the data in Table 6, then we obtain $\hat{\alpha}_{2.} < 0$ and $\hat{\beta}_{.3} < 0$, that is, boundary solutions occur. Now we need to verify if both Conditions 1 and 2 of Theorem 5.2 hold. For the matrix A^T and the vector $\mathbf{b}$, we have

$$456 < a_{12} \times \frac{b_2}{a_{22}} + a_{13} \times \frac{b_3}{a_{33}} = 260 \times \frac{156}{131} + 93 \times \frac{125}{18} = 955.4516,$$

$$156 < a_{21} \times \frac{b_1}{a_{11}} + a_{23} \times \frac{b_3}{a_{33}} = 290 \times \frac{456}{621} + 30 \times \frac{125}{18} = 421.2802,$$

$$125 < a_{31} \times \frac{b_1}{a_{11}} + a_{32} \times \frac{b_2}{a_{22}} = 284 \times \frac{456}{621} + 117 \times \frac{156}{131} = 347.8693,$$

so that Condition 1 in Theorem 5.2 is satisfied. Also, for the matrix A and the vector $\mathbf{b}^*$, we have

$$135 < a_{12} \times \frac{b_2^*}{a_{22}} + a_{13} \times \frac{b_3^*}{a_{33}} = 290 \times \frac{60}{131} + 284 \times \frac{20}{18} = 448.38,$$

$$60 < a_{21} \times \frac{b_1^*}{a_{11}} + a_{23} \times \frac{b_3^*}{a_{33}} = 260 \times \frac{135}{621} + 117 \times \frac{20}{18} = 186.5217,$$

$$20 < a_{31} \times \frac{b_1^*}{a_{11}} + a_{32} \times \frac{b_2^*}{a_{22}} = 93 \times \frac{135}{621} + 30 \times \frac{60}{131} = 33.9578,$$

so that Condition 2 in Theorem 5.2 is satisfied. Further, from Table 7, we observe that boundary solutions also occur if Models [M1]-[M4] are fitted to data in Table 6. Then only Condition 1 is satisfied if boundary solutions under [M1] and [M3] occur, while only Condition 2 is satisfied if boundary solutions under [M2] and [M4] occur. This is because the MLE $\hat{\alpha} = (\hat{\alpha}_{i.})$ under Models [M1] and [M3] satisfies the system $A^T \alpha = \mathbf{b}$, while the MLE $\hat{\beta} = (\hat{\beta}_{.j})$ under Models [M2] and [M4] satisfies the system $A\beta = \mathbf{b}^*$.

5.4. The necessary conditions are not sufficient

The next example shows that the necessary conditions for the occurrence of boundary solutions in Theorem 5.2 are not sufficient.

Example 5.3. In Example 5.2, replace 456 by 366 in $\mathbf{b}$ and 20 by 15 in $\mathbf{b}^*$ so that $\mathbf{b} = (366, 156, 125)$ and $\mathbf{b}^* = (135, 60, 15)$ now. For the matrix A^T and the vector $\mathbf{b}$, we have

$$366 < a_{12} \times \frac{b_2}{a_{22}} + a_{13} \times \frac{b_3}{a_{33}} = 260 \times \frac{156}{131} + 93 \times \frac{125}{18} = 955.4516,$$

$$156 < a_{21} \times \frac{b_1}{a_{11}} + a_{23} \times \frac{b_3}{a_{33}} = 290 \times \frac{366}{621} + 30 \times \frac{125}{18} = 379.2512,$$

$$125 < a_{31} \times \frac{b_1}{a_{11}} + a_{32} \times \frac{b_2}{a_{22}} = 284 \times \frac{366}{621} + 117 \times \frac{156}{131} = 306.7099,$$

so that Condition 1 in Theorem 5.2 is satisfied. Also, for the matrix A and the vector $\mathbf{b}^*$, we have

$$135 < a_{12} \times \frac{b_2^*}{a_{22}} + a_{13} \times \frac{b_3^*}{a_{33}} = 290 \times \frac{60}{131} + 284 \times \frac{15}{18} = 369.4911,$$

$$60 < a_{21} \times \frac{b_1^*}{a_{11}} + a_{23} \times \frac{b_3^*}{a_{33}} = 260 \times \frac{135}{621} + 117 \times \frac{15}{18} = 154.0217,$$

$$15 < a_{31} \times \frac{b_1^*}{a_{11}} + a_{32} \times \frac{b_2^*}{a_{22}} = 93 \times \frac{135}{621} + 30 \times \frac{60}{131} = 33.9578,$$

so that Condition 2 in Theorem 5.2 is satisfied. Now, when we solve the system $A^T \alpha = \mathbf{b}$, then we obtain the MLE's $\hat{\alpha}_{1.} = 0.0133$, $\hat{\alpha}_{2.} = 0.7796$ and $\hat{\alpha}_{3.} = 1.6671$. So, there are no boundary solutions under Model [M3]. Similarly, the system $A\beta = \mathbf{b}^*$ yields the MLE's $\hat{\beta}_{.1} = 0.041$, $\hat{\beta}_{.2} = 0.3655$ and $\hat{\beta}_{.3} = 0.0126$, that is, there are no boundary solutions under Model [M4]. Since the MLE's in Model [M5] satisfy both the systems $A^T \alpha = \mathbf{b}$ and $A\beta = \mathbf{b}^*$, there are no boundary solutions under [M5] as well. Similar results hold for Models [M1] and [M2]. Hence, the conditions in Theorem 5.2 are not sufficient for the occurrence of boundary solutions under Models [M1]-[M5].

5.5. Importance of the necessary conditions

Here, we discuss additional details about Theorem 5.2 and discuss its simplicity and effectiveness.

From Theorem 5.2, note that if $\{y_{i+12}\}$, $\{y_{+j21}\}$, and/or $\{\hat{\mu}_{ii11}\}$ are large, then Conditions 1 and 2 may not hold. Indeed, if the inequalities in Conditions 1 and 2 are reversed for all $1 \leq i \leq I$ and $1 \leq j \leq I$, then from statements (a),

(b) and (c) of Theorem 5.2, boundary solutions do not occur on fitting Models [M1]-[M5] in an $I \times I \times 2 \times 2$ incomplete table.

It is known that when boundary solutions occur, perfect fit models (here Models M3], [M4] and [M5]) cannot reproduce the observed counts, indicating poor fit and imprecision of the parameter estimates. The MLE's of the parameters under NMAR models lie on the boundary of the parameter space and the log likelihood function tends to be flat, which makes derivation of the MLE's computationally intensive. Also, the corresponding covariance matrix has unreasonable eigenvalues (close to either zero or negative), which implies the estimated standard errors for some parameter estimates are large. Hence, for model selection, we prefer NMAR models which don't yield boundary solutions upon fitting them to the given data.

Theorem 5.1 provides conditions, which help us identify the occurrence of boundary solutions. However, boundary solutions may occur under some NMAR models if any of the sufficient conditions in Theorem 5.1 does not hold. This implies that Theorem 5.1 cannot always provide us the set of plausible NMAR models for model selection. However, note that Theorem 5.2 is very useful in this regard since it gives us an insight into verifying the non-occurrence of boundary solutions under each of the NMAR models [M1]-[M5]. That is, if any of the necessary conditions in Theorem 5.2 does not hold, then we know for sure that boundary solutions do not occur. This always helps us to obtain the list of candidate NMAR models suitable for fitting the given data. Hence, Theorem 5.2 is more reliable than Theorem 5.1 for the purpose of model selection in square two-way incomplete tables.

The non-boundary MLE's of μ_{ij11} are $\hat{\mu}_{ij11} = \frac{y_{ij11}y_{i+1+}y_{++11}}{y_{i+11}y_{++1+}}$ under Model [M1], $\hat{\mu}_{ij11} = \frac{y_{ij11}y_{+j+1}y_{++11}}{y_{+j11}y_{+++1}}$ under Model [M2], and $\hat{\mu}_{ij11} = y_{ij11}$ under Models [M3], [M4] and [M5] (see pp. 647-648 of [2]), which involve only the observed cell counts and their sums. Hence, from Theorem 5.2, there is no need to solve any system of likelihood equations, use the EM algorithm or compute odds (based on the observed (joint/marginal) cell counts) to check for the non-occurrence of boundary solutions in an $I \times I \times 2 \times 2$ incomplete table.

Remark 5.1. If $A_D = \text{diag}(a_{11}, \ldots, a_{II})$, then from [6], the solutions $\alpha = (\alpha_{i.})$ of the system $A^T \alpha = \mathbf{b}$ may be obtained iteratively as follows.

$$\alpha^{(0)} = A_D^{-1}\mathbf{b}$$

(5.4) $$\alpha^{(n+1)} = \alpha^{(n)} + A_D^{-1}(\mathbf{b} - A^T\alpha^{(n)}), \quad n = 0, 1, 2, \ldots.$$

Similarly, the solutions $\beta = (\beta_{.j})$ of the system $A\beta = \mathbf{b}^*$ may be obtained iteratively as follows.

$$\beta^{(0)} = A_D^{-1}\mathbf{b}^*$$

(5.5) $$\beta^{(n+1)} = \beta^{(n)} + A_D^{-1}(\mathbf{b}^* - A\beta^{(n)}), \quad n = 0, 1, 2, \ldots.$$

Both the sequences (5.4) and (5.5) converge to the solutions of the respective systems.

6.   CONCLUSIONS

In this paper, we have discussed the problem of boundary solutions that occur under various NMAR models for an $I \times J \times 2 \times 2$ table. We formally define boundary solutions for such a table and provide a result (Proposition 3.1) that theoretically connects and justifies various forms of these solutions under alternative parametrizations of the missing data models. This eliminates the need of using the EM algorithm (see pp. 39-40 of [9]) to empirically obtain the forms of the solutions in two-way incomplete tables. The above result is then used to improve a claim in [2] regarding the occurrence of boundary solutions. We give the precise forms of such solutions by just noting the corresponding level (s) of the variable (s) in the table, which reduces computational burden.

As discussed earlier, boundary solutions pose a lot of problems for estimation and inference under NMAR models in incomplete tables. Hence, it is important to investigate sufficient and necessary conditions for their occurrence in such tables. We have provided a result (Theorem 5.1) on sufficient conditions for the occurrence of boundary solutions in an $I \times J \times 2 \times 2$ table. While [9] consider only Model [M5], we consider Models [M1]-[M5] in Theorem 5.1. We use a similar approach but give direct arguments instead of contrapositive ones used in Theorem 1 of [9] for proving Theorem 5.1. [7] conjectured that these conditions would also be necessary for general two-way incomplete tables. However, we show by a counterexample that this is not the case for $I, J \geq 3$, thereby disproving the conjecture.

We have also established necessary conditions in Theorem 5.2 for the occurrence of boundary solutions in an $I \times J \times 2 \times 2$ table, which have not been discussed in the literature so far. As discussed in Section 5.5, these conditions are of practical utility to identify the non-occurrence of boundary solutions and hence for model selection. However, we show by a counterexample that these conditions are not sufficient. Note that a major advantage of the proposed sufficient conditions and necessary conditions is that they depend only on the observed cell counts in the table or their sums. As mentioned in [9], this makes the verification process much easier, and avoids using the EM algorithm or solving likelihood equations. Finally, all the above results are illustrated using six data analysis examples. It would be helpful to obtain a set of conditions involving only the observed cell counts, which are sufficient as well as necessary for the occurrence of boundary solutions in two-way incomplete tables with both variables missing.

ACKNOWLEDGMENTS

2/98 (SA-I). The authors are also grateful to the referees for carefully reading the manuscript and suggesting numerous improvements.

Appendix

A1

Proof of Proposition 3.1: From Definition 3.1, it follows that if boundary solutions occur under the Models [M1]-[M5], then the MLE's of the cell probabilities except some of the nonresponse ones are all non-zero. On substituting $k = l = 1$ (for response cell probabilities) in the above models and using the parameter constraints, we can then deduce that the MLE's of the constant, the main effects and the association terms between Y_i's, between R_i's, and between Y_i and R_j for $i \neq j$ are all finite. This is because non-zero terms (response cell probabilities) on the LHS of the log-linear models imply that the log-linear parameters on the RHS are finite.

Consider part 1 first. For the Models [M1] and [M3], the log-linear parameters modelling the non-ignorable nonresponse (NMAR) mechanism of Y_1 are $\lambda_{R_1}(k)$ and $\lambda_{Y_1 R_1}(i, k)$. If boundary solutions occur, then they are of the form $\hat{\pi}_{ij2l} = 0$ (see point 1 of Definition 3.1), which implies $\hat{\lambda}_{Y_1 R_1}(i, 2) = -\infty$ for at least one i since the other parameters are finite as mentioned above. Then under Model [M1], we have

$$
\begin{aligned}
\hat{\pi}_{i+2+} &= \sum_{j,l} \hat{\pi}_{ij2l} \\
&= \frac{1}{N} \sum_{j,l} \exp\{\hat{\lambda} + \hat{\lambda}_{Y_1}(i) + \hat{\lambda}_{Y_2}(j) + \hat{\lambda}_{R_1}(2) + \hat{\lambda}_{R_2}(l) + \hat{\lambda}_{Y_1 R_1}(i, 2) \\
&\quad + \hat{\lambda}_{Y_1 Y_2}(i, j) + \hat{\lambda}_{R_1 R_2}(2, l)\} \\
&= 0
\end{aligned}
$$

for at least one i. Conversely, we have

$$
\begin{aligned}
&\hat{\pi}_{i+2+} = 0 \quad \text{(for at least one } i) \\
&\Rightarrow \sum_{j,l} \exp\{\hat{\lambda} + \hat{\lambda}_{Y_1}(i) + \hat{\lambda}_{Y_2}(j) + \hat{\lambda}_{R_1}(2) + \hat{\lambda}_{R_2}(l) + \hat{\lambda}_{Y_1 R_1}(i, 2) + \hat{\lambda}_{Y_1 Y_2}(i, j) \\
&\quad + \hat{\lambda}_{R_1 R_2}(2, l)\} = 0 \\
&\Rightarrow \hat{\lambda}_{Y_1 R_1}(i, 2) = -\infty \quad \text{for at least one } i,
\end{aligned}
$$

so that $\hat{\lambda}_{Y_1 R_1}(i, 2) = -\infty \Leftrightarrow \hat{\pi}_{i+2+} = 0$ for at least one i under Model [M1]. The same can be shown for Model [M3]. Under Models [M1] and [M3], $a_{ij} =$

$\exp[2\{\lambda_{R_1}(2) + \lambda_{Y_1 R_1}(i,2) + \lambda_{R_1 R_2}(2,1)\}]$. Since a_{ij} depends only on i, we have $a_{ij} = \alpha_{i.}$. It is clear that $\hat{\alpha}_{i.} = 0 \Leftrightarrow \hat{\lambda}_{Y_1 R_1}(i,2) = -\infty$. Also, note that by definition of a_{ij}, if $\hat{\alpha}_{i.} = 0$ for all $1 \le i \le I$, then $y_{+j21} = 0$ for all $1 \le j \le J$, which is a contradiction since supplementary margins are assumed to be positive. Hence, under Models [M1] and [M3], boundary solutions are given by $\hat{\lambda}_{Y_1 R_1}(i,2) = -\infty \Leftrightarrow \hat{\pi}_{i+2+} = 0 \Leftrightarrow \hat{\alpha}_{i.} = 0$ for at least one and at most $(I-1)$ values of Y_1.

Consider part 2 now. Under Models [M2] and [M4], the log-linear parameters modelling the NMAR nechanism of Y_2 are $\lambda_{R_2}(l)$ and $\lambda_{Y_2 R_2}(j,l)$. Also, $b_{ij} = \exp[2\{\lambda_{R_2}(2) + \lambda_{Y_2 R_2}(j,2) + \lambda_{R_1 R_2}(1,2)\}]$. Since b_{ij} depends only on j, we have $b_{ij} = \beta_{.j}$. Then it can be shown similarly as above that boundary solutions in this case are given by $\hat{\lambda}_{Y_2 R_2}(j,2) = -\infty \Leftrightarrow \hat{\pi}_{+j+2} = 0 \Leftrightarrow \hat{\beta}_{.j} = 0$ for at least one and at most $(J-1)$ values of Y_2.

Finally, consider part 3. Under Model [M5], the log-linear parameters modelling the NMAR nechanisms of Y_1 and Y_2 are $\lambda_{R_1}(k)$, $\lambda_{R_2}(l)$, $\lambda_{Y_1 R_1}(i,k)$ and $\lambda_{Y_2 R_2}(j,l)$. The proof for the form of boundary solutions under Model [M5] follows on similar lines as for Models [M1]-[M4] shown above.

A2

Proof of Theorem 5.1: From [2], the MLE's $\hat{\alpha}_{i.}$ under the NMAR model for only Y_1 (Models [M1] and [M3]) satisfy

(6.1) $$\sum_i N\hat{\pi}_{ij11}\hat{\alpha}_{i.} = y_{+j21}, \quad \forall\, 1 \le j \le I,$$

while the MLE's $\hat{\beta}_{.j}$ under the NMAR model for only Y_2 (Models [M2] and [M4]) satisfy

(6.2) $$\sum_j N\hat{\pi}_{ij11}\hat{\beta}_{.j} = y_{i+12}, \quad \forall\, 1 \le i \le I.$$

The MLE's $\hat{\alpha}_{i.}$ and $\hat{\beta}_{.j}$ under the NMAR model for both Y_1 and Y_2 (Model [M5]) satisfy both (6.1) and (6.2). Note that boundary solutions in Models [M1] and [M3] occur if $\hat{\alpha}_{i.} \le 0$ for at least one and at most $(I-1)$ values of Y_1, while boundary solutions in Models [M2] and [M4] occur if $\hat{\beta}_{.j} \le 0$ for at least one and at most $(I-1)$ values of Y_2. Also note that boundary solutions under [M5] occur if at least one of the following holds:

(i) $\hat{\alpha}_{i.} \le 0$ for at least one and at most $(I-1)$ values of Y_1,

(ii) $\hat{\beta}_{.j} \le 0$ for at least one and at most $(I-1)$ values of Y_2.

From (5.1) and (6.1), we have

$$\nu(j,j') = \frac{y_{+j21}}{y_{+j'21}} = \frac{\sum_i \hat{\pi}_{ij11}\hat{\alpha}_{i.}}{\sum_i \hat{\pi}_{ij'11}\hat{\alpha}_{i.}},$$

(6.3) $$\nu_m(j,j') - \nu(j,j') = \frac{\sum_{i \neq m_1}(\hat{\pi}_{m_1j11}\hat{\pi}_{ij'11} - \hat{\pi}_{m_1j'11}\hat{\pi}_{ij11})\hat{\alpha}_{i.}}{\hat{\pi}_{m_1j'11}\sum_i \hat{\pi}_{ij'11}\hat{\alpha}_{i.}},$$

(6.4) $$\nu(j,j') - \nu_n(j,j') = \frac{\sum_{i \neq n_1}(\hat{\pi}_{n_1j'11}\hat{\pi}_{ij11} - \hat{\pi}_{n_1j11}\hat{\pi}_{ij'11})\hat{\alpha}_{i.}}{\hat{\pi}_{n_1j'11}\sum_i \hat{\pi}_{ij'11}\hat{\alpha}_{i.}},$$

where m_1 and n_1 are the levels of Y_1 corresponding to $\nu_m(j,j')$ and $\nu_n(j,j')$ respectively. From (5.1), we get

(6.5) $$\nu_n(j,j') = \frac{\hat{\pi}_{n_1j11}}{\hat{\pi}_{n_1j'11}} < \nu_i(j,j') = \frac{\hat{\pi}_{ij11}}{\hat{\pi}_{ij'11}} < \nu_m(j,j') = \frac{\hat{\pi}_{m_1j11}}{\hat{\pi}_{m_1j'11}}.$$

From (6.5), we have the following inequalities

(6.6) $$\hat{\pi}_{m_1j11}\hat{\pi}_{ij'11} > \hat{\pi}_{m_1j'11}\hat{\pi}_{ij11}, \ \hat{\pi}_{n_1j'11}\hat{\pi}_{ij11} > \hat{\pi}_{n_1j11}\hat{\pi}_{ij'11} \ \text{for } i \neq m_1, n_1.$$

Consider part (a). Suppose Condition 1 holds, which implies that (6.3) and (6.4) are of opposite signs. Using this fact and (6.6), we observe that $\hat{\alpha}_{i.} < 0$ for at least one and at most $(I-1)$ values of Y_1, that is, boundary solutions of the form $\hat{\pi}_{i+2+} = 0$ occur.

Again from (5.2) and (6.2), we have

$$\omega(i,i') = \frac{y_{i+12}}{y_{i'+12}} = \frac{\sum_j \hat{\pi}_{ij11}\hat{\beta}_{.j}}{\sum_j \hat{\pi}_{i'j11}\hat{\beta}_{.j}},$$

(6.7) $$\omega_m(i,i') - \omega(i,i') = \frac{\sum_{j \neq m_2}(\hat{\pi}_{im211}\hat{\pi}_{i'j11} - \hat{\pi}_{i'm211}\hat{\pi}_{ij11})\hat{\beta}_{.j}}{\hat{\pi}_{i'm211}\sum_i \hat{\pi}_{i'j11}\hat{\beta}_{.j}},$$

(6.8) $$\omega(i,i') - \omega_n(i,i') = \frac{\sum_{j \neq n_2}(\hat{\pi}_{i'n211}\hat{\pi}_{ij11} - \hat{\pi}_{in211}\hat{\pi}_{i'j11})\hat{\beta}_{.j}}{\hat{\pi}_{i'n211}\sum_i \hat{\pi}_{i'j11}\hat{\beta}_{.j}},$$

where m_2 and n_2 are the levels of Y_2 corresponding to $\omega_m(i,i')$ and $\omega_n(i,i')$ respectively. From (5.2), we get

(6.9) $$\omega_n(i,i') = \frac{\hat{\pi}_{in211}}{\hat{\pi}_{i'n211}} < \omega_j(i,i') = \frac{\hat{\pi}_{ij11}}{\hat{\pi}_{i'j11}} < \omega_m(i,i') = \frac{\hat{\pi}_{im211}}{\hat{\pi}_{i'm211}}.$$

From (6.9), we have the following inequalities

(6.10) $$\hat{\pi}_{m_2j11}\hat{\pi}_{ij'11} > \hat{\pi}_{m_2j'11}\hat{\pi}_{ij11}, \ \hat{\pi}_{n_2j'11}\hat{\pi}_{ij11} > \hat{\pi}_{n_2j11}\hat{\pi}_{ij'11} \ \text{for } j \neq m_2, n_2.$$

Now consider part (b). Assume Condition 2 holds, which implies that (6.7) and (6.8) are of opposite signs. Using this fact and (6.10), we observe that $\hat{\beta}_{\cdot j} < 0$ for at least one and at most $(I - 1)$ values of Y_2, that is, boundary solutions of the form $\hat{\pi}_{+j+2} = 0$ occur.

Finally consider part (c). Assume at least one of Conditions 1 and 2 holds. The cases when only Condition 1 holds or only Condition 2 holds follow from the proofs of part (a) and part (b) respectively. So it is sufficient here to assume both Conditions 1 and 2 hold. This implies, from part (a), $\hat{\alpha}_{i\cdot} < 0$ for at least one and at most $(I - 1)$ values of Y_1, that is, boundary solutions of the form $\hat{\pi}_{i+2+} = 0$ occur. Also from part (b), we have $\hat{\beta}_{\cdot j} < 0$ for at least one and at most $(I - 1)$ values of Y_2, that is, boundary solutions of the form $\hat{\pi}_{+j+2} = 0$ occur. This completes the proof.

A3

Proof of Theorem 5.2: From Theorem 5.1, the MLE's $\hat{\alpha}_{i\cdot}$ and $\hat{\beta}_{\cdot j}$ under Model [M5] satisfy

$$(6.11) \qquad \sum_i \hat{\mu}_{ij11}\hat{\alpha}_{i\cdot} = y_{+j21} \quad \text{for } j = 1, \ldots, I,$$

$$(6.12) \qquad \sum_j \hat{\mu}_{ij11}\hat{\beta}_{\cdot j} = y_{i+12} \quad \text{for } i = 1, \ldots, I.$$

Also, the MLE $\hat{\alpha}_{i\cdot}$ under Models [M1] and [M3] satisfy (6.11) only, while the MLE $\hat{\beta}_{\cdot j}$ under Models [M2] and [M4] satisfy (6.12) only. Note that boundary solutions under [M5] occur if at least one of the following conditions hold:

(i) $\hat{\alpha}_{i\cdot} \leq 0$ for at least one and at most $(I - 1)$ values of Y_1,

(ii) $\hat{\beta}_{\cdot j} \leq 0$ for at least one and at most $(I - 1)$ values of Y_2.

Also, boundary solutions in Models [M1] and [M3] are given by only Condition (i), while boundary solutions in Models [M2] and [M4] are given by only Condition (ii). In Lemma 5.1, take $A = (\hat{\mu}_{ij11})$, $\mathbf{b} = (b_j) = (y_{+j21})$ and $\mathbf{b}^* = (b_i^*) = (y_{i+12})$ for $1 \leq i \leq I$, $1 \leq j \leq I$. Then (6.11) may be written as $A^T\alpha = \mathbf{b}$, while (6.12) may be written as $A\beta = \mathbf{b}^*$, where $\alpha = (\alpha_{i\cdot})$ and $\beta = (\beta_{\cdot j})$. We prove Theorem 5.2 by contrapositive.

Consider part (a) first. Suppose Condition 1 in Theorem 5.2 does not hold. Then by Lemma 5.1, $\alpha = (A^T)^{-1}\mathbf{b} > \mathbf{0}$. In other words, $\hat{\alpha}_{i\cdot} > 0$ for all $1 \leq i \leq I$, that is, boundary solutions under Models [M1] and [M3] do not occur.

Consider part (b) now. Assume Condition 2 in Theorem 5.2 does not hold. Then by Lemma 5.1, $\beta = A^{-1}\mathbf{b}^* > \mathbf{0}$. In other words, $\hat{\beta}_{.j} > 0$ for all $1 \leq j \leq I$, that is, boundary solutions under Models [M2] and [M4] do not occur.

Finally consider part (c). Assume both Conditions 1 and 2 in Theorem 5.2 do not hold. Then by Lemma 5.1, both $\hat{\alpha}_{i.} > 0$ and $\hat{\beta}_{.j} > 0$ for all $1 \leq i \leq I$, $1 \leq j \leq I$, that is, boundary solutions under Model [M5] do not occur.

Hence, the result follows.

REFERENCES

[1]   BAKER, S. G. AND LAIRD, N. M. (1988). Regression analysis for categorical variables with outcome subject to nonignorable nonresponse, *Journal of the American Statistical Association*, **83**, 62-69.

[2]   BAKER, S. G.; ROSENBERGER, W. F. AND DERSIMONIAN, R. (1992). Closed-form estimates for missing counts in two-way contingency tables, *Statistics in Medicine*, **11**, 643-657.

[3]   CLARKE, P. S. (2002). On boundary solutions and identifiability in categorical regression with non-ignorable non-response, *Biometrical Journal*, **44**, 701-717.

[4]   CLARKE, P. S. AND SMITH, P. W. F. (2005). On maximum likelihood estimation for log-linear models with non-ignorable non-responses, *Statistics and Probability Letters*, **73**, 441-448.

[5]   GHOSH, S. AND VELLAISAMY, P. (2016). On the occurrence of boundary solutions in multidimensional incomplete tables, *Statistics and Probability Letters*, **119**, 63-75.

[6]   KAYKOBAD, M. (1985). Positive Solutions of Positive Linear Systems, *Linear Algebra and its Applications*, **64**, 133-140.

[7]   KIM, S. AND PARK, Y. (2014). Power-linear models for incomplete contingency tables with nonignorable non-responses, *Statistics*, DOI: 10.1080/02331888.2013.869595.

[8]   LITTLE, J. A. AND RUBIN, D. B. (2002). *Statistical Analysis with Missing Data*, second ed., Wiley, New York.

[9]   PARK, Y.; KIM, D. AND KIM, S. (2014). Identification of the occurrence of boundary solutions in a contingency table with nonignorable nonresponse, *Statistics and Probability Letters*, **93**, 34-40.

[10]   SMITH, P. W. F.; SKINNER, C. J. AND CLARKE, P. S. (1999). Allowing for non-ignorable nonresponse in the analysis of voting intention data, *Journal of the Royal Statistical Society: Series C*, **48**, 563-577.